Original Research Article

MDM Medical Decision Making

Medical Decision Making
2022, Vol. 42(1) 94–104

DOI: 10.1177/0272989X211011160
journals.sagepub.com/home/mdm
SAGE

Effect of Having, but Not Consulting, a Computerized Diagnostic Aid

Mark V. Pezzo, Brenton E. D. Nash, Pierre Vieux, and Hannah W. Foster-Grammer

Previous research has described physicians' reluctance to use computerized diagnostic aids (CDAs) but has never experimentally examined the effects of not consulting an aid that was readily available. **Experiment 1.** Participants read about a diagnosis made either by a physician or an auto mechanic (to control for perceived expertise). Half read that a CDA was available but never actually consulted; no mention of a CDA was made for the remaining half. For the physician, failure to consult the CDA had no significant effect on competence ratings for either the positive or negative outcome. For the auto mechanic, failure to consult the CDA actually increased competence ratings following a negative but not a positive outcome. Negligence judgments were greater for the mechanic than for the physician overall. **Experiment 2.** Using only a negative outcome, we included 2 different reasons for not consulting the aid and provided accuracy information highlighting the superiority of the CDA over the physician. In neither condition was the physician rated lower than when no aid was mentioned. Ratings were lower when the physician did not trust the CDA and, surprisingly, higher when the physician believed he or she already knew what the CDA would say. Finally, consistent with our previous research, ratings were also high when the physician consulted and then followed the advice of a CDA and low when the CDA was consulted but ignored. Individual differences in numeracy did not qualify these results. Implications for the literature on algorithm aversion and clinical practice are discussed.

Keywords
algorithm aversion, artificial intelligence, computerized decision support systems, decision aids, individual differences, myth of expertise, numeracy

Date received: April 27, 2020; accepted: March 26, 2021

Reducing medical errors is an important challenge for health care. Diagnostic errors, in particular, account for 7% to 17% of adverse hospital outcomes and 10% of patient deaths[1] and are the leading cause of paid malpractice claims.[2] Many such errors can be reduced by the use of computerized diagnostic aids (CDAs). Extensive research has shown that the actuarial approach used by CDAs is almost always equal to or better than clinical judgment alone.[3,4] As a result, the use of CDAs typically improves clinical outcomes.[5–8] Nevertheless, physicians are sometimes reluctant or even hostile about using CDAs.[9–15]

Resistance can be caused by a number of factors,[16,17] including unfamiliarity, difficulty using the technology,[1] and overconfidence/repudiation.[18,19] Some physicians may fear that consulting an aid will undermine their credibility and expertise.[20,21] The myth of physician expertise is particularly strong,[10,22] sometimes resulting in the derogation of physicians who receive "help" from an aid.[23,24] Nevertheless, most patients realize that physicians make mistakes.[25] Do patients really want doctors to ignore potentially valuable diagnostic tools? In this article, we report the effect of choosing not to consult an available decision aid on patients' perceptions of physician competence and negligence. In addition, we replicate and extend our earlier research in which a CDA is consulted and its advice is either ignored or followed.[26]

Corresponding Author
Mark V. Pezzo, University of South Florida, 140 7th Ave South, DAV 113, St. Petersburg, FL 33701, USA (pezzo@usf.edu).

Patient Attitudes toward Medical Decision Aids

The first researchers to experimentally examine how patients perceive physicians who use CDAs were Pezzo and Pezzo.[26] As compared with a no-aid control condition, when the physician followed the advice of a CDA, he or she was rated less negatively following a bad outcome but less positively following a good outcome. Furthermore, when a missed diagnosis occurred because a radiologist followed a CDA's recommendation, competence ratings did not suffer significantly, particularly if he or she agreed with the CDA. However, when the physician consulted but ignored the CDA's advice, competence ratings were significantly lower although not lower than when a CDA was never mentioned. Thus, using a CDA can sometimes protect the physician by being assigned some of the responsibility for a bad outcome.

Arkes et al.[27] asked participants to view a mock malpractice trial concerning an unnecessary appendectomy. The decision aid was not computerized but rather a hand-calculated Alvarado Score that recommended appendectomy over a specific threshold. Use of this score had no effect on judgments of liability. However, if the physician was considered liable, he was scrutinized more severely if he had consulted the aid but defied it than if he followed its recommendation.

Additional research, however, produced divergent results.[23] As compared with a physician who made a diagnosis unassisted, one who used a computer program was judged as having inferior diagnostic skills in 3 separate scenarios. Individual differences in participants' numeracy skills did not qualify these findings. This research was later replicated by Wolf and colleagues using both information technology students and Indian phsyicians.[28,29]

Shaffer et al.[24] wondered whether the participants in these previous studies were more upset because the physician needed help in making a diagnosis (the myth of expertise) or because using a computer somehow dehumanizes the doctor–patient relationship. A physician decided whether to order a diagnostic test for an ankle injury either 1) alone, 2) after consulting a CDA, or 3) after consulting another expert (human). Judgments made alone or after consulting an expert produced higher physician ratings than judgments made after consulting the CDA. A second study found that positive attitudes toward statistics predicted greater satisfaction with the doctor but did not moderate the effects of consulting an aid.

Nonmedical Perceptions of Decision Aids

The accounting profession has begun to rely on CDAs, particularly when the perceived risk of litigation is high. In one study, practicing auditors judged a hypothetical auditor less harshly following a mistake if he or she had relied entirely, rather than partially, on a CDA.[30] Notably, however, using a CDA actually lowered auditors' confidence in their own decision, presumably because many auditors are not aware of the high accuracy of CDAs, simply using them to avoid litigation rather than inform an optimal decision.[31]

There is considerable resistance toward decision aid use in hiring decisions.[10] Diab et al.[32] found that when participants were told that a mathematical formula was used to select which job applicant to employ, they found it less useful, less legal, and more unprofessional than when the applications were "examined and thoroughly discussed" in a holistic fashion. This is consistent with Arkes et al.[23] who found that physicians who used aids were judged to be less competent, professional, and thorough than those who did not.

An Unfair Comparison?

There appears to be an important difference between studies that demonstrate a protective effect of CDAs and those that do not. Studies reporting physician derogation following the use of a decision aid did not provide outcome information,[23,24,28,29] whereas participants in studies showing a protective effect of aids did.[26,27,33,34] We suspect that not providing an outcome exaggerates the degree to which participants perceive the CDA as a liability. Because judgments of negligence are rarely made in the absence of outcome information, studies including outcomes may better approximate the real-world consequences of using a CDA or not.

Previous researchers have also used scenarios describing somewhat simple aids. For example, Arkes et al.[23] and Shaffer et al.[24] used a simple decision rule called the Ottawa Ankle Rule (OAR), which specifies that an x-ray series is required only to evaluate fractures if there is pain in specific areas of the ankle or foot. Participants who read the scenario, however, might wonder why a computer is needed at all. The OAR does not use a complex algorithm, and many patients assume that an x-ray alone is all a physician needs to make a determination. In fact, efforts to increase education regarding the OAR have not been shown to reduce unnecessary trips to the hospital.[35] It makes sense, then, that participants might question why a physician would

University of South Florida, St. Petersburg, FL, USA (MVP, BED, PV); Claremont Graduate University, Upland, CA, USA (HWF-G).
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors received no financial support for the research, authorship, and/or publication of this article.

need a CDA to make such an "easy" decision, despite the fact that it is a helpful rule that reduces errors.[36]

The Unused Decision Aid

Past research has included conditions in which a CDA was never mentioned, was consulted and heeded, and was consulted but then ignored. To our knowledge, no medical decision-making studies have presented information about an available CDA that was never consulted. Such research has been conducted, however, in forensic auditing.

In one study,[34] participants were told an auditor either did not consult, consulted, but then ignored, or heeded a decision aid said to have low (81%) or high (90%) accuracy. Auditors who consulted but ignored the high accuracy aid were deemed most responsible for an audit failure. Those who did not have or did not consult the aid were rated equally less responsible. Finally, auditors who heeded the aid were judged least responsible. No conditions differed for the low-accuracy aid.

A second study[33] varied the extent to which input from CDA was used, including a condition in which an aid was available but completely ignored. Any use of the decision aid reduced both the auditor's judged responsibility for a financial loss and the amount of damages he or she awarded. Interestingly, the degree of reliance made no difference in judged responsibility.[37]

Why would consulting but ignoring the advice of a CDA differ from failing to consult it at all? It may depend on the degree of counterfactual processing that each produces. Arguably, it is easier to undo a negative outcome when the doctor sought—but ignored—potentially useful information than when no such information was sought. Although not as strong, this same logic also suggests that not being aware of a CDA at all should result in less blame than being told that one existed but was not consulted.

The Present Research

In experiment 1, we informed half of the participants that a CDA was available to the decision maker but never actually consulted. A CDA was never mentioned to the remaining half. Because perceived expertise might qualify the findings, half of the participants read about a physician, whereas the other half read about an auto mechanic. The mechanic scenario was developed to parallel the physician scenario as closely as possible. We chose an auto mechanic because the diagnostic task is similar to that of physicians, because mechanics have diagnostic aids at their disposal, and because mechanics require considerably less education and presumed expertise than do physicians.[38] We predicted that both the mechanic and physician would be judged more harshly when choosing not to consult a CDA but that the effect would be weaker for the physician, who will, presumably, be believed by some not to need a CDA.

Experiment 1

Method

Participants. Students at the University of South Florida St. Petersburg ($N = 225$) received course credit for participating (female = 72%; $M_{age} = 28.9$ years; $Mdn = 25$). The ethnicity of the sample reflected the US population overall (White = 71.4%; Black = 9.8%; Hispanic = 9.4%; Asian = 2.7%; other = 6.7%).

Procedure and measures. Individuals were randomly assigned to read 1 of 8 scenarios in a 2 (decision aid) × 2 (outcome) × 2 (profession) factorial design. Participants were asked to imagine seeking the diagnostic help of either a doctor for chest pain or an auto mechanic for their car's erratic performance. The diagnosis was said to result in either a negative or a positive outcome. Half the participants were told the decision maker had a CDA available but never actually consulted it. The remaining half were not informed of any CDA. Participants were then asked to list thoughts describing how they "would feel if [they] were actually in this situation" and then, as a manipulation check, to rate the "overall quality of the outcome" (1 = *very bad*; 5 = *very good*). Finally, participants rated the extent to which the decision maker was competent, responsible for the outcome, negligent, and deserves to be sued (1 = *strongly disagree*; 5 = *strongly agree*). See Supplementary Appendix A and B for complete materials.

Results

Manipulation check. We excluded data from participants who, inexplicably, rated the negative outcome as "good" or "very good" ($n = 5$), or rated the positive outcome as "bad" or "very bad" ($n = 9$), or failed to provide an outcome quality rating ($n = 4$). The 207 remaining participants rated the positive outcome much higher ($M = 4.17$, $SD = 0.79$) than the negative condition ($M = 1.73$, $SD = 0.73$), $F(1, 203) = 573.3$, $P < 0.001$, $\eta^2_{partial} = 0.74$. A small main effect of profession indicated that both outcomes were judged more positive for the mechanic ($M = 3.07$, $SD = 1.50$) than for the

Table 1 Competence and Negligence Judgments as a Function of Outcome, Decision Aid, and Profession (Experiment 1)

Source	*df*	*F*	*P*	η_p^2
DV = competence				
Outcome	1	64.3	<0.001	0.244
Aid	1	2.67	0.104	0.013
Profession	1	1.59	0.209	0.008
Aid × Profession	1	1.01	0.315	0.005
Aid × Outcome	1	1.18	0.279	0.006
Profession × Outcome	1	18.88	<0.001	0.087
Aid × Profession × Outcome	1	6.21	0.014	0.030
Error	199			
DV = negligence				
Aid	1	1.34	0.25	0.007
Profession	1	12.01	0.001	0.056
Aid × Profession	1	<1	0.94	<0.001
Error	203			

Note. Cell sizes ranged from $n = 22$ to 32 (competence) and from $n = 46$ to 57 (negligence). Negligence is a composite of 3 measures: negligent, liability, and responsibility (Cronbach's $\alpha = 0.81$) and includes only participants who received a negative outcome.

physician ($M = 2.87$, $SD = 1.39$), $F(1, 203) = 11.86$, $P = 0.002$, $\eta^2_{\text{partial}} = 0.06$. There was no interaction, $F(1, 203) = 2.6$, $P = 0.11$.

Experience. Eleven participants (14%) reported some experience with the medical profession (e.g., nurse, emergency medical technician [EMT]) or evaluation of the profession (e.g., law firm). An additional 11 participants (12.5%) had at least semiprofessional experience working in the auto mechanic profession. These participants were included in all analyses reported here; analyses excluding them left the results (significance and effect sizes) largely unchanged.

Competence. Judgments of competence were submitted to a 2 (outcome) × 2 (decision aid) × 2 (profession) analysis of variance (ANOVA). Most important was a 3-way interaction (see Table 1), which can be interpreted as follows: of 4 possible comparisons, in only 1 did not consulting the aid affect competence ratings as compared with the control condition (Ps > 0.18 for the other 3 comparisons). Specifically, and contrary to our prediction, when the mechanic did not consult the CDA in the negative outcome condition, competence ratings were higher than when no CDA was mentioned (see Figure 1; $P = 0.005$, $\eta^2_{\text{partial}} = 0.79$). In addition, results showed a large main effect of outcome, with positive outcomes resulting in higher competence ratings ($M = 3.98$, $SD = 0.82$) than negative outcomes ($M = 2.95$, $SD = 1.05$). This was qualified by the profession × outcome interaction. The simple main effects of outcome were significant for both professions (Ps ≤ 0.01), but the effect of outcome was much stronger for the mechanic ($\eta^2_{\text{partial}} = 0.275$) than for the physician ($\eta^2_{\text{partial}} = 0.033$). There was no main effect of profession, and neither the decision aid × profession nor the decision aid × outcome interactions were significant.

Negligence. Ratings of negligence, responsibility, and liability following a negative outcome were combined into a single composite variable called "negligence," with good reliability (Cronbach's $\alpha = 0.81$). Composite negligence ratings were submitted to 2 a (decision aid) × 2 (profession) ANOVA. Outcome was excluded as a factor, because these measures were not taken in the positive condition. Values for F, df, P, and effect size may be found in Table 1 (bottom). Negligence was judged somewhat higher overall for the mechanic ($M = 3.91$, $SD = 0.91$) than for the physician ($M = 3.42$, $SD = 0.89$), but there was no significant effect of CDA, nor was there a significant interaction between decision aid and profession. Compared with the control condition, neither profession was judged more negligent when they failed to consult the CDA.

Thought listings. Analyses of participant thoughts can be found in the supplemental materials.

Discussion

Competence. As expected, decision makers were rated as less competent following a negative than a positive outcome. This effect was larger for the mechanic than it was for the physician. More important, not consulting the

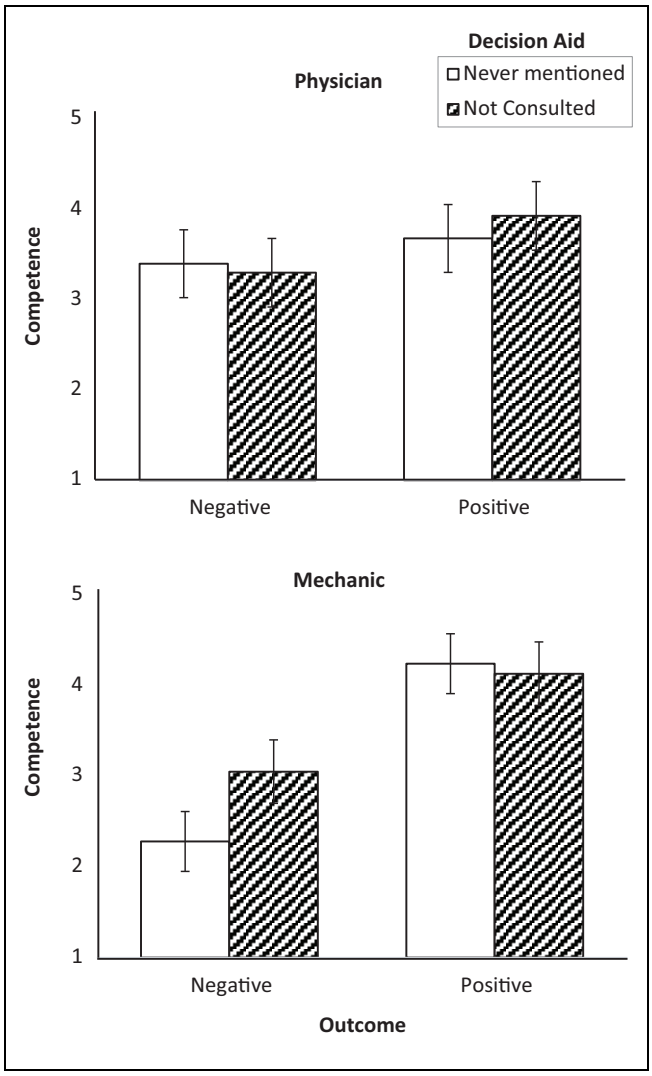

Figure 1 Competence judgments as a function of outcome, decision aid, and profession (experiment 1). A negative outcome resulted in lower competence ratings than did positive outcomes, and this effect was stronger for the mechanic than for the physician. There was no overall effect of consulting the aid, but a 3-way interaction showed that for the mechanic (but not the physician), there was a benefit to not consulting the computerized diagnostic aid in the negative condition. Error bars denote 95% confidence interval.

aid had no effect on competence ratings for the physician in either the positive or negative condition nor for the mechanic in the positive condition (see Figure 1). In the positive condition, this may reflect a feeling that CDA consultation was not needed. Indeed, 35% of participants in the positive condition explicitly stated that the physician probably did not need to use the aid, and 55% did so for the mechanic (see supplemental analyses). This does not explain the lack of effect for the physician in the negative condition, however, as only 5% of participants here claimed the CDA was unnecessary. Finally, and surprisingly, competence ratings increased when the mechanic chose not to consult the CDA following a negative outcome. One possible explanation is that the mere mention of a decision aid focuses attention on the difficulty of the diagnostic task, buffering the mechanic from lower competence judgments.

Negligence. Composite judgments of negligence (combining ratings of negligence, responsibility, and liability) were taken only following a negative outcome. Surprisingly, the only significant finding was that, overall, mechanics were judged as more negligent than physicians for the missed diagnosis. This might reflect an overall distrust of mechanics independent of CDA use. Indeed, following a positive outcome, many more respondents expressed pleasure and surprise that their mechanic was honest/capable (see supplemental analyses). This supports the idea that physicians, even when they make mistakes, are able to maintain their reputation more so than are auto mechanics. More importantly, independent of profession, those who did not consult an aid were not rated any more negligent than those who did not have an aid available.

Decision Aid Accuracy

Experiment 1 did not provide accuracy information concerning the unused CDA. Past research has rarely included such information. One exception[26] provided accuracy rates for both CDA (93%) and the typical physician (84%) and reported a benefit of using the aid. A second described "a computer program developed at the prestigious Mayo Clinic, one of the nation's premier medical facilities", but found no consistently beneficial effect of using the aid.[23(p198)] Carlson and others[39] reported that of 29 potential factors hypothesized to produce trust in an aid, "reputation of the manufacturer" ranked near the bottom. In contrast, "statistics of the machine's past performance" ranked very high in determining patient trust in an aid. Indeed, recent research indicates that patients are willing to see an artificially intelligent provider if the artificial intelligence (AI) is more accurate than the human provider is.[40,41] In the absence of accuracy information, lay people most likely assume that a physician is more accurate than a CDA.[9] This was confirmed by our thought listing analysis, at least following a positive outcome. The reality is that CDAs, particularly those using AI, now regularly outperform human experts in a variety of diagnostic tasks.[42–46]

Studies that exclude accuracy information may unfairly bias their findings against CDA acceptance. Thus, in experiment 2, we include this information.

Experiment 2

Experiment 1 showed that having but not consulting a CDA did not result in lower competence ratings or in greater negligence judgments. In experiment 2, we replicate and extend these findings with a new scenario and additional conditions in which the aid is consulted and then either followed or not. To better understand the findings of experiment 1, we also included 2 different reasons for not consulting the aid. The diagnostic accuracy of both the aid and a typical physician were included. Finally, we measured individual differences in numeracy[47] to examine whether participants' basic understanding of probability might qualify the findings. To increase the power of the analyses, experiment 2 dropped both the positive outcome and mechanic conditions. All participants received a negative outcome.

Method

Participants. Students at the University of South Florida (N = 235) received course credit for participating (female = 75%; M_{age} = 26.8 years; Mdn = 23). Racial makeup roughly paralleled the US makeup (White = 57.9%; Black = 16.2%; Hispanic = 14.0%; Asian = 2.6%; other/missing = 9.3%). Twenty-one participants said that they had some experience in the medical profession (e.g,. EMT, nurse, etc.).

Scenario. A patient sought medical attention after experiencing a sudden, blinding headache. The physician diagnosed migraine headache and sent the patient home. Two weeks later, the patient suffered a rupture of a cerebral aneurysm with complications, and it was determined that the headache was "sentinel" (i.e., a sign of an impending aneurysm). The doctor was sued and asked during the trial how the diagnosis was made. Participants were then randomly assigned to 1 of 6 possible conditions:

1. *Control*: Aid never mentioned
2. *Defy*: Aid consulted; physician disagreed with diagnosis and ignored it
3. *No Consult 1*: Aid available; not consulted; physician does not trust it
4. *No Consult 2*: Aid available; not consulted; physician believed he knew what it would say
5. *Heed*: Aid consulted; physician initially disagrees, but ultimately heeds advice
6. *Agree*: Aid consulted; physician agrees with diagnosis

In all but the control condition, participants were told the CDA had 93% accuracy, and the typical physician had 79% accuracy. See Supplementary Appendix C and D for complete materials.

Measures. As a manipulation check, participants indicated how *bad* the outcome was. They then rated how *obvious, foreseeable*, and *surprising* (reverse coded) they found the outcome, the extent to which the physician *should have known* that the medical condition was more serious than a migraine, and the *likelihood* of the negative outcome (0%–100%). The physician was then rated on 4 measures: responsibility, competence, negligence, and should be sued. Participants reported gender, age, race, and medical experience and then completed the abbreviated numeracy scale.[47]

Results

Participants perceived the outcome as very bad (M = 4.31, SD = 0.74). This did not differ across conditions, $F(5, 229) < 1$, P = 0.93. Participants also did not differ across conditions in terms of how surprising, obvious, foreseeable, or likely the outcome was (all Ps > 0.07), and these ratings were all close to the midpoint of their respective scale. These measures were taken to test additional hypotheses and will not be discussed further. A detailed account, however, may be found in Pezzo et al.[48]

Because the outcome was always negative, we combined all 4 primary measures into a composite score. We first reverse scored "responsible," "negligent," and "should be sued" so that higher values corresponded to more favorable physician ratings. Next, we converted each of the 4 ratings into z-scores and took their mean to produce the final composite physician rating (Cronbach's α = 0.76). A 1-way ANOVA indicated a marginally significant effect of condition on the physician rating, $F(5, 229)$ = 2.17, P = 0.058, $\eta^2_{partial}$ = 0.05.

As the primary goal of this research was to compare all conditions in which the physician has a CDA available to the condition wherein it was never mentioned, we analyzed 5 planned contrasts. As shown in Table 2, 3 of the conditions (No Consult 2, Heed, and Agree), produced physician ratings significantly greater than in the control, albeit with relatively small effect sizes (ds = 0.26 to 0.28). The Defy and No Consult 1 conditions,

Table 2 Planned Contrasts Comparing Mean Composite Rating for Control Condition to All Remaining Conditions (Experiment 2)

Condition	*n*	M_{comp}	M_{diff}	*P*	*d*
Control	38	–0.1868	—	—	—
Defy	41	–0.1873	–0.0005	0.998	<0.001
No Consult 1 (Trust)	40	–0.0939	0.0929	0.586	0.07
No Consult 2 (Know)	39	0.1578	0.3446	0.045	0.27
Heed	39	0.1518	0.3386	0.049	0.26
Agree	38	0.1732	0.3600	0.038	0.28

Note. M_{comp} = composite physician rating (*z* score); M_{diff} = M_{comp}– control.

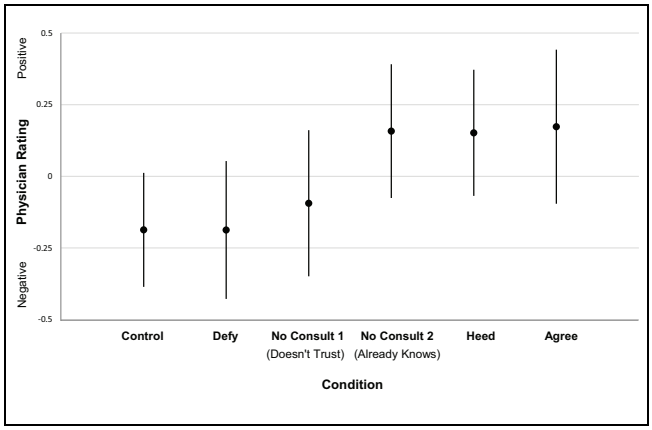

Figure 2 Composite physician rating as a function of computerized diagnostic aid condition (experiment 2).

however, did not differ from the control condition. The overall pattern of results can be seen in Figure 2 and was similar for all variables contributing to the composite.

Eleven participants did not complete the numeracy scale; $N = 224$ remained for analysis. Numeracy scores ($\alpha = 0.68$) were not significantly correlated with the composite physician rating, $r(224) = 0.12$, $P = 0.08$, nor did they significantly qualify the effect of condition in predicting physician rating, $P = 0.12$. Additional analyses can be found in the supplemental materials.

Discussion

Experiment 2 described a physician who had a CDA that was said to outperform most physicians. As long as the physician consulted and followed the advice of the aid, the CDA seemed to protect the physician. Despite a misdiagnosis that resulted in an aneurysm, the physician was still rated more positively than when no CDA was mentioned. When the aid was consulted, but then ignored, however, the physician was rated lower, but no more so than when the aid was never mentioned. In this study, individual differences in numeracy[47] did not directly affect physician ratings nor did they qualify the effects of condition on those ratings.

Finally, consider what happened when the physician never consulted the CDA. In the first condition (No Consult 1), the physician did not consult the aid because he did not trust it. Here, as expected, the physician was rated relatively low, although again, not lower than in the control condition.

It could be that the presence of a CDA reminds participants of how difficult such diagnoses are and serves to elevate their ratings of the physician—even when the aid is ignored, not consulted at all!

In the second condition (No Consult 2), the physician did not consult the aid because they assumed the aid would come to the same conclusion as they did. Remarkably, the physician was rated significantly higher than in the control condition and virtually identical to the agree and heed conditions. We discuss possible reasons and implications for this below.

General Discussion

Physicians can be resistant to seeking advice from a CDA.[9–15] Our primary goal in this research was to examine what happens when a physician refuses to consult a CDA that is readily available. Surprisingly, the effects were not particularly negative. In experiment 1, refusing to consult a CDA did not affect either negligence or competence judgments; they were no lower than a control condition in which no CDA was ever mentioned. To explore why this might be, we included 2 different reasons for not consulting the aid in experiment 2. We also provided accuracy information to show the general superiority of the CDA over the typical physician.[40,41] When the physician's reason was a distrust of the CDA, ratings were again no different than in the no-CDA control condition. Participants appeared to believe—despite

being informed of the CDA's superiority—that consulting it would have done little to prevent the negative outcome. Had they believed it would be helpful, we would expect physician ratings to fall below those in the no-CDA condition.

Remarkably, when the physician claimed to already know what the CDA would say, participants rated him significantly more positive than they did in the no-aid (and distrust) conditions and as positive as when the CDA was actually consulted and its advice followed. We assumed that such a claim would appear somewhat arrogant, especially in light of a missed diagnosis. Instead, participants gave the physician the benefit of the doubt, consistent with the myth of physician expertise.[10,22] Although participants never received the foregone advice from the unconsulted CDA, it appears they believed the advice to be consistent with the physician's own diagnosis. Had they believed it to be inconsistent, participants likely would have engaged in counterfactual "second guessing" and given lower ratings, similar to what occurred in the (logically equivalent) consult-but-defy condition.

A secondary goal of this research was to replicate the findings of previous research showing a protective effect of consulting a CDA.[26,27,33,34] Indeed, we found that when a misdiagnosis results in a negative outcome, having followed the advice of the CDA can actually protect physicians against increased negligence judgments. Our replication used a novel scenario and confirmed that the protection afforded physicians occurs regardless of whether they originally agreed or disagreed with the CDA. The physician who consulted but then ignored (i.e., defy) the aid's advice, however, was judged more negatively. Also consistent with past research, neither attitudes toward statistics[24] nor numeracy[23] qualified the effects of CDA use on physician ratings in this research.

Algorithm Aversion

Our research intersects with recent studies on algorithm aversion.[49] Consistent with this literature, failure to consult a CDA sometimes mitigated physician derogation following a negative outcome. However, consulting and then following the CDA's advice equally reduced physician derogation. Clearly, algorithm aversion is not ubiquitous.[41] One reason for the latter finding may be that we informed participants that the CDA was more accurate than the typical physician. Recent research shows that people will choose an artificially intelligent medical provider over a human physician when the AI is known to be superior.[40] An equally important cause of algorithm aversion, however, is that the underlying process is often not transparent.[50] Even though a "black box" algorithm may have been demonstrated to be accurate in the past, many humans still find it difficult to trust something that they cannot explain.[51] Eliminating algorithm aversion, then, will likely depend on our ability to develop "explainable" AI.[52,53]

Limitations

Our findings may not generalize to all populations. Both studies were conducted at a single university with disproportionate gender representation. The use of hypothetical scenarios may also have affected our results, although similar results have been found using additional medical scenarios[26] and a realistic video presentation.[27] Our expertise manipulation in experiment 1 was confounded with profession. Although this allowed us to generalize beyond physicians, future research should compare expert versus nonexpert decision makers within a single profession. A reviewer predicted that senior physicians may be judged more harshly, presumably because they should know to always consult the CDA. Finally, participants in experiment 2 had to take the physician's word that the unconsulted CDA advice was predictable. This could be verified easily in a trial setting by simply comparing the physician's diagnosis to that of the CDA.

Future Research

This research shows that the reason for choosing not to consult a CDA matters to potential jurors. Future research should examine this further. We suspect that a physician who claims to already know what the CDA will advise appears more knowledgeable and confident than does a physician who doesn't trust the CDA. Research on nonphysicians has found that highly confident (even overconfident) individuals are typically believed to be more skilled and more deserving of respect and admiration than equally skilled individuals who display less confidence.[54] Overconfidence has also been linked to undesirable behaviors among health professionals,[55,56] including a resistance to admitting to medical mistakes.[57] Ironically, when combined with an apology, admitting to a medical mistake has been shown to reduce the likelihood of malpractice lawsuits.[58]

Implications for Clinical Practice and Policy

This research found that physicians may sometimes receive more blame if they consult but do not follow the advice of a CDA than if they do not consult it at all. Of

course, we do not recommend that physicians avoid consulting CDAs, nor do we recommend that physicians blindly rely on CDAs simply to avoid lawsuits.[31,59] CDAs are not perfect, and physicians should not be punished for ignoring advice they deem to be incorrect. Unfortunately, very little research has focused on the cognitive and metacognitive demands placed on physicians required to use sophisticated systems.[60] Such demands may make it difficult for physicians to catch mistakes made by the AI, particularly those physicians with limited experience or who work under time pressure.[61] One possible solution to the human–CDA interface problem is better regulation. While it is true that CDAs typically outperform humans in controlled research studies,[5–8] there has been enormous growth of companies producing CDAs (particularly AI based), and the barriers to entry for such companies are quite low.[62] As decision aids are highly specialized, federal regulations are unlikely to solve the problem. Indeed, regulation of CDAs in both the United States and Europe is inconsistent, with relatively few agreed-upon standards. Instead, professional medical organizations should actively seek to certify CDAs to ensure not only their accuracy but also transparency and ease of use.[63]

One important form of transparency involves the reporting of how the algorithm was trained and tested, since algorithms that were trained on only a subset of the population may be effective only on that subset.[64] Perhaps more important, transparency requires information about the underlying process used by the algorithm to make a recommendation. Such information may sometimes be difficult to obtain, however, either because it is proprietary or is simply unknown, as is the case with deep-learning AI. As a result, some researchers suggest that we abandon such black box approaches altogether in favor of much simpler and completely transparent fast-and-frugal decision aids.[51] Regardless of the underlying complexity involved, CDAs should be designed for humans.[65] Human-centered design[66] can help to create CDAs that provide information quickly and concisely in a way that matches the physician's diagnostic workflow.[67] Ultimately, CDAs should be designed so that physicians want to use them and not feel obligated to do so.

Acknowledgments

We thank Jason Beckstead, Samantha Boddy, Bharat Bharat, and Tara Brooks for comments on an earlier version of this article. We thank Russell Upchurch for his help in coding the data and Stephanie Taylor for her assistance in developing the scenario used in experiment 2.

ORCID iDs

Mark V. Pezzo https://orcid.org/0000-0002-4442-6244
Hannah W. Foster-Grammer https://orcid.org/0000-0002-5724-7579

Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* website at https://journals.sagepub.com/home/mdm.

References

1. Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, The National Academies of Sciences, Engineering, and Medicine. *Improving Diagnosis in Health Care*. Balogh EP, Miller BT, Ball JR, eds. Washington (DC): National Academies Press (US); 2016.
2. Tehrani ASS, Lee H, Mathews SC, et al. 25-year summary of US malpractice claims for diagnostic errors 1986–2010: an analysis from the National Practitioner Data Bank. *BMJ Qual Saf*. 2013;22(8):672–80.
3. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science*. 1989;243(4899):1668–74.
4. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. *Psychol Assess*. 2000;12(1):19–30.
5. Garg AX, Adhikari NKJ, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA*. 2005;293(10):1223–38.
6. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems. *Ann Intern Med*. 2012;157:29.
7. Lynch CJ, Liston C. New machine-learning technologies for computer-aided diagnosis. *Nat Med*. 2018;24(9):1304–5.
8. Berner ES, La Lande TJ. Overview of clinical decision support systems. In: Berner ES, ed. *Clinical Decision Support Systems: Theory and Practice*. New York: Springer; 2007. p. 3–22.
9. Arkes HR. Being an advocate for linear models of judgment is not an easy life. In: Dawes RM, Krueger JI, eds. *Rationality and Social Responsibility: Essays in Honor of Robyn Mason Dawes*. Boca Raton (FL): CRC Press; 2008. p. 47–70.
10. Highhouse S. Stubborn reliance on intuition and subjectivity in employee selection. *Ind Organ Psychol*. 2008;1(3):333–42.
11. Lai F, Macmillan J, Daudelin DH, Kent DM. The potential of training to increase acceptance and use of computerized decision support systems for medical diagnosis. *Hum Factors*. 2006;48(1):95–108.
12. Freudenheim M. Many hospitals resist computerized patient care. *New York Times*. 2004;6. Available from: https://www.andrew.cmu.edu/user/rk2x/mmmjune05/itheal th.pdf

13. Coiera E. Technology, cognition and error. *BMJ Qual Saf*. 2015;24(7):417–22.
14. Eccles M, McColl E, Steen N, et al. Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial. *BMJ*. 2002;325(7370):941.
15. 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.
16. Litvin CB, Ornstein SM, Wessell AM, Nemeth LS, Nietert PJ. Adoption of a clinical decision support system to promote judicious use of antibiotics for acute respiratory infections in primary care. *Int J Med Inform*. 2012;81(8):521–6.
17. Liberati EG, Ruggiero F, Galuppo L, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. *Implement Sci*. 2017;12(1):113.
18. Sieck WR, Arkes HR. The recalcitrance of overconfidence and its contribution to decision aid neglect. *J Behav Decis Mak*. 2005;18(1):29–53.
19. Arkes HR, Dawes RM, Christensen C. Factors influencing the use of a decision rule in a probabilistic task. *Organ Behav Hum Decis Process*. 1986;37:93–110.
20. Kaplan B. Evaluating informatics applications—clinical decision support systems literature review. *Int J Med Inform*. 2001;64(1):15–37.
21. Liang H, Xue Y. Face Loss and resistance to clinical decision support systems. 2016. Available from: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1004&context=sighci2016
22. Dawes RM. *House of Cards: Psychology and Psychotherapy Built on Myth*. New York: Free Press; 1994.
23. Arkes HR, Shaffer VA, Medow MA. Patients derogate physicians who use a computer-assisted diagnostic aid. *Med Decis Making*. 2007;27(2):189–202.
24. Shaffer VA, Probst CA, Merkle EC, Arkes HR, Medow MA. Why do patients derogate physicians who use a computer-based diagnostic support system? *Med Decis Making*. 2013;33(1):108–18.
25. Gallagher TH, Waterman AD, Ebers AG, Fraser VJ, Levinson W. Patients' and physicians' attitudes regarding the disclosure of medical errors. *JAMA*. 2003;289(8):1001–7.
26. 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(1):48–56.
27. 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(2):201–8.
28. Wolf JR. *Do IT Students Prefer Doctors Who Use IT?* Working Paper. Normal: Illinois State University; 2014. p. 1–34.
29. Joshi V, Wolf J. Do Indian physicians derogate peers who use computer-based diagnostic aids? In: *Conference Presentation. AMCIS 2011 Proceedings*. 2011. p. Paper 298.
30. Lowe DJ, Reckers P. The influence of outcome effects, decision aid usage, and intolerance of ambiguity on evaluations of professional audit judgement. *Int J Accounting Audit Peformance Eval*. 1997;1(December 1995):43–58.
31. Gomaa MI, Hunton JE, Rose JM. The effect of control risk and litigation risk on decision aid reliance. *Int J Appl Decis Sci*. 2008;1(1):80–106.
32. Diab DL, Pui SY, Yankelevich M, Highhouse S. Lay perceptions of selection decision aids in US and non-US samples. *Int J Select Assess*. 2011;19(2):209–16.
33. Chan SH, Lowe DJ, Yao LJ. The legal implications of auditors using a fraud decision aid vs. professional judgment. *J Forens Account*. 2008;9(1):63–82.
34. Lowe DJ, Reckers P, Whitecotton SM. The effects of decision-aid use on jurors' evaluations of auditor liability. *Rev Acct Stud*. 2002;77(1):185–202.
35. Blackham J, Claridge T, Benger JR. Can patients apply the Ottawa ankle rules to themselves? *Emerg Med*. 2008;25(11):750–1.
36. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Ann Emerg Med*. 1992;21(4):384–90.
37. Anderson JC, Jennings MM, Kaplan SE, Reckers PMJ. The effect of using diagnostic decision aids for analytical procedures on judges' liability judgments. *J Account Public Policy*. 1995;14(1):33–62.
38. Chory-Assad RM, Tamborini R. Television exposure and the public's perceptions of physicians. *J Broadcast Electron Media*. 2003;47(2):197–215.
39. Carlson MS, Desai M, Drury JL, Kwak H, Yanco HA. Identifying factors that influence trust in automated cars and medical diagnosis systems. In: *The Intersection of Robust Intelligence and Trust in Autonomous Systems: Papers from the AAAI Spring Symposium*. 2011. p. 20–7.
40. Pezzo MV, Beckstead JW. Patients prefer artificial intelligence to a human provider, provided the AI is better than the human: a commentary on Longoni, Bonezzi and Morewedge (2019). *Judgm Decis Making*. 2020;15(3):443.
41. Pezzo MV, Beckstead JW. Algorithm aversion is too often presented as though it were non-compensatory: a reply to Longoni et al. (2020). *Judgm Decis Mak*. 2020;15(3):449–51.
42. Esteva A, Kuprel B, Novoa RA, et al. Corrigendum: dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;546(7660):686.
43. Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*. 2016;529(7587):484–9.
44. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*. 2017;12(4):e0174944.
45. Strickland E. IBM Watson, heal thyself: how IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*. 2019;56(4):24–31.

46. Yeomans M, Shah A, Mullainathan S, Kleinberg J. Making sense of recommendations. *J Behav Decis Mak*. 2019;29:84.
47. Weller JA, Dieckmann NF, Tusler M, Mertz CK, Burns WJ, Peters E. Development and testing of an abbreviated numeracy scale: a Rasch analysis approach. *J Behav Decis Mak*. 2013;26(2):198–212.
48. Pezzo MV, Quinn MS, Machacek MG, Marchal C. *Does Hindsight Bias Mediate Outcome Bias in Judgments of Decision Quality and Negligence?* Working Paper. University of South Florida St. Petersburg; 2020.
49. Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *J Exp Psychol Gen*. 2015;144(1):114–26.
50. Marcus G, Davis E. *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Knopf Doubleday; 2019.
51. Katsikopoulos KV, Simsek O, Buckmann M, Gigerenzer G. *Classification in the Wild: The Science and Art of Transparent Decision Making*. Cambridge (MA): MIT Press; 2021.
52. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion*. 2020;58:82–115.
53. Rudin C, Radin J. Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review*. 2019;1(2). Available from: https://hdsr.mitpress.mit.edu/pub/f9kuryi8
54. Kennedy JA, Anderson C, Moore DA. When overconfidence is revealed to others: testing the status-enhancement theory of overconfidence. *Organ Behav Hum Decis Process*. 2013;122(2):266–79.
55. Lima de Miranda K, Detlefsen L, Stolpe M. *Overconfidence and Hygiene Non-compliance in Hospitals*. Kiel Working Paper; 2020. Report No.: 2156. Available from: https://www.econstor.eu/handle/10419/222296
56. Meyer AND, Payne VL, Meeks DW, Rao R, Singh H. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. *JAMA Intern Med*. 2013; 173(21):1952–8.
57. Brezis M, Orkin-Bedolach Y, Fink D, Kiderman A. Does physician's training induce overconfidence that hampers disclosing errors? *J Patient Saf*. 2019;15(4):296–8.
58. Ho B, Liu E. Does sorry work? The impact of apology laws on medical malpractice. *J Risk Uncertain*. 2011;43(2):141.
59. Hamilton JG, Genoff Garzon M, Westerman JS, et al. "A tool, not a crutch": patient perspectives about IBM Watson for oncology trained by Memorial Sloan Kettering. *J Oncol Pract*. 2019;JOP1800417.
60. Jussupow E, Spohrer K, Heinzl A, Gawlitza J. Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Inform Syst Res*. 2021. Available from: https://doi.org/10.1287/isre.2020.0980
61. Gomaa M, Gomaa A. The effect of time pressure, task complexity, and litigation risk on auditors reliance on decision aids. In: Proceedings of the Hawai'i Accounting Research Conference. Shidler College, University of Hawaii at Manoa; January 3–5, 2019. p. 1–33.
62. Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. *Science*. 2019;363(6429): 810–2.
63. Price WN II, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. *JAMA*. 2019. Available from: http://dx.doi.org/10.1001/jama.2019.15064
64. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. *J Am Med Inform Assoc*. 2020;27(12):2011–5.
65. Bates DW, Kuperman GJ, Wang S, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc*. 2003;10(6):523–30.
66. Babione JN, Ocampo W, Haubrich S, et al. Human-centred design processes for clinical decision support: a pulmonary embolism case study. *Int J Med Inform*. 2020;142: 104196.
67. Chan TM, Mercuri M, Turcotte M, Gardiner E, Sherbino J, de Wit K. Making decisions in the era of the clinical decision rule: how emergency physicians use clinical decision rules. *Acad Med*. 2020;95(8):1230–7.