Physician Evaluation after Medical Errors: Does Having a Computer Decision Aid Help or Hurt in Hindsight?

Mark V. Pezzo, PhD, Stephanie P. Pezzo, BA

Objective. *The authors examined whether physicians' use of computerized decision aids affects patient satisfaction and/or blame for medical outcomes.* ***Method.*** *Experiment 1: Fifty-nine undergraduates read about a doctor who made either a correct or incorrect diagnosis and either used a decision aid or did not. All rated the quality of the doctor's decision and the likelihood of recommending the doctor. Those receiving a negative outcome also rated negligence and likelihood of suing. Experiment 2: One hundred sixty-six medical students and 154 undergraduates read negative-outcome scenarios in which a doctor either agreed with the aid, heeded the aid against his own opinion, defied the aid in favor of his own opinion, or did not use a decision aid. Subjects rated doctor fault and competence and the appropriateness of using decision aids in medicine. Medical students made judgments for themselves and for a layperson.* ***Results.*** *Experiment 1: Using a decision aid caused a positive outcome to be rated less positively and a negative outcome to be rated less negatively. Experiment 2: Agreeing with or heeding the aid was associated with reduced fault, whereas defying the aid was associated with roughly the same fault as not using one at all. Medical students were less harsh than undergraduates but accurately predicted undergraduate's responses.* ***Conclusion.*** *Agreeing with or heeding a decision aid, but not defying it, may reduce liability after an error. However, using an aid may reduce favorability after a positive outcome.* ***Key words:*** *computer-assisted diagnosis; decision support systems; clinical; decision support techniques; diagnostic errors; malpractice; medical errors; medical informatics; patient satisfaction.* ***(Med Decis Making 2006;26:48–56)***

A widely cited Institute of Medicine report declared that between 44 000 and 98 000 hospital deaths per year occurred as a result of medical errors.[1] Although the exact numbers are debated,[2,3] a common suggestion for reducing errors has been the implementation of computerized decision aids. Such aids typically combine large databases of base rate information with mathematical modeling to arrive at probabilistic estimates for different possible outcomes. The seminal research on this technology found that aid-based diagnoses for abdominal pain were superior to unaided clinician's diagnoses using the same patient information.[4] In the 3 decades since this research, decision aids have been used in a variety of areas, most notably the monitoring of physician (drug) orders to prevent dangerous drug interactions.[5] A systematic review found that in 43 of 65 studies, there was at least some benefit in either the process of care or patient outcomes.[6]

Clearly, decision aids can help physicians. However, although attitudes about the decision aids among physicians are generally positive,[7] the use of such devices is still not widespread.[8,9] One recent study reported that 4th-year medical undergraduates used decision aids infrequently in clinical settings.[10] Students' attitudes in this study were significantly affected by faculty attitudes toward such devices, but virtually no research has examined physician attitudes toward decision aids.

Received 28 February 2005 from Psychological Sciences Program, University of South Florida, St. Petersburg, FL (MVP); College of Medicine, University of South Florida, Tampa, FL (SPP). We thank Hal Arkes and Jason Beckstead for helpful comments on an earlier draft of this article and Ronda Keefe for her assistance in data coding and entry. Revision accepted for publication 3 August 2005.

Address correspondence to Mark V. Pezzo, 140 7th Ave, South, DAV 258, University of South Florida, St. Petersburg, FL 33701; e-mail: pezzo@stpt.usf.edu.

DOI: 10.1177/0272989X05282644

One exception is a brief report by Grundmeier and Johnson.[11] House staff physicians at 2 academic medical centers who did not regularly use decision aids read scenarios about the use of an aid for a surgical consult and as a hypokalemia alert. In general, 63% of house staff physicians were in favor of a decision aid if it had a positive predictive value above 67%. The majority of respondents who gave written comments were in favor of decision aids primarily because they believed they would improve quality of care. However, the authors noted that this small sample was biased by "an apparent over representation of computer-savvy housestaff."[11(p269)] Of the physicians who did offer concerns, a common complaint, seen in 13% of written comments, was that "physicians won't think for themselves."[11(p269)] Most other complaints centered on concerns that decision aids would be annoying, inconvenient, and slow and that too many messages would lead to their being ignored.

RELUCTANCE TO USE DECISION AIDS

Decision makers have historically refused to rely on decision aids. Boatsman and others[12] found that auditors ignored fraud predictions of a decision aid in 67% of cases. Ashton[13] reported that only 2 of 91 relied completely on an available decision aid. Arkes and others[14] reported a reliance rate of 56%, and Powell[15] found a reliance rate of 52%. A survey of 2 decades' worth of medical decision aids use showed considerably lower reliance rates, with no medical journals reporting any more than 32% reliance, and most showing considerably less.[16]

People seem particularly reluctant to use decision aids that consist of a single rule or simple statistical formula. Arkes and others[14] found that self-professed sports experts refused to use a simple decision rule, even when informed that the rule would outperform their predictions. Presumably, the experts believed, despite being told otherwise, that they could outperform the decision aid, particularly because the aid relied on fewer cues than were available.[17] Even if it can be shown that such a rule outperforms human decision makers, people still fear that it may miss important "broken leg cues"[18] that would render the rule irrelevant in a particular instance. Indeed, Ashton[13] found that auditors rely less on decision aids that use fewer cues, irrespective of their predictive accuracy. Ashton also showed that increased pressure to perform well actually decreased reliance on a decision aid.

Besides feeling pressure to outperform the good but not perfect performance guaranteed by the aid, decision makers may also be concerned about making justifiable decisions. Decision makers might worry that their decisions would be harder to justify if they simply accepted the recommendations of the decision aid or contradicted the aid's advice. Boatsman and others[12] suggest strategic or intentional nonreliance is common, driven primarily by concerns about the consequences of reliance (e.g., justification pressures) rather than by concerns about the accuracy of the decision aid. Conversely, there has been speculation that after a poor outcome, a physician could be held liable for having failed to use a decision aid, even if it is not the standard of care.[19] However, very few studies have examined the effect of ignoring a decision aid on judgments of fault or perceived competence of the decision maker, and none have examined these factors in the context of medical decision making.

RESEARCH IN AUDITING

Although no research has directly addressed the effects of ignoring diagnostic decision aids on judgments of medical decision quality, some relevant literature exists in the auditing field. Decision aids are primarily used to help auditors detect fraud. As with doctors, auditors seem to be hesitant to use decision aids because of fears that failure to follow the aid might result in greater culpability. However, only a handful of studies have evaluated whether these fears are warranted. Lowe and Reckers[20] asked subjects to evaluate a supervisor who had erred after either fully using a decision aid (investigating all 15 possibilities that it generated) or partially using the aid (investigating only 10 possibilities). Results indicated that judgments were significantly more harsh when subjects believed that the supervisor had only partially used the decision aid. In a later study, Lowe and others[21] found that auditors who heeded an incorrect decision aid were blamed less for a bad outcome than were those who had no decision aid available to them (controls). Furthermore, those who had a decision aid but ignored its recommendations were blamed more for the outcome than were controls.

Overview

We present 2 studies designed to assess the effect of using or ignoring the diagnostic advice of a decision aid to assess whether such aids are beneficial or detrimental when physicians' decisions are evaluated in hindsight. For example, is a doctor who heeds the advice of a decision aid considered less negligent after an error than the doctor who did not use an aid because

the aid is partially responsible for the decision? Or, is the doctor perceived as less competent for needing to enlist the advice of a computer decision aid in the first place? Alternatively, doctors who use such modern technology may be perceived as better trained and more cutting edge after a positive outcome. In Experiment 1, we examine these possibilities in the case of a doctor who either heeds the advice of the decision aid or uses no aid at all, resulting in either a positive or a negative outcome. In Experiment 2, we restrict our focus to negative outcomes, paying specific attention to the consequences of consulting but then disagreeing with the diagnosis provided by a decision aid. To examine the extent to which concerns about increased blame in this situation might contribute to doctors' reluctance to embrace diagnostic decision aids, Experiment 2 also assesses medical students' assumptions about laypeople's opinions of decision aids.

EXPERIMENT 1

Method

Subjects

Fifty-nine Introductory Psychology students at a selective university received course credit for participating. Students at this university were required to own a laptop computer and were generally considered more sophisticated and accepting of technology than was the general population.

Scenario

All subjects were given a medical scenario and asked to respond "as though you were really in this situation." All subjects were to imagine that they were young (late 20s) and athletic, with a family history of heart disease, and that they had recently checked into the hospital after experiencing intermittent but severe chest pain for 4 days. A doctor considers several possible causes (e.g., gas, inflammation of the chest wall, heart disease) and orders a series of tests. Subjects in the positive outcome condition read that the doctor was able to detect a severe but easily correctable condition, and thus they would enjoy a complete recovery. Subjects in the negative outcome condition read that the doctor missed the correctable condition, leading to considerable decrease in their quality of life.

Computer Condition

Half of all subjects were told that the doctor used a computer decision aid that determined the most likely diagnosis by combining all test results according to an empirically derived mathematical formula. Subjects were informed that the computer has been shown to be correct about 93% of the time, whereas doctors are correct on average only about 84% of the time. The doctor then stated that "because the computer generally predicts better than doctors," he heeded the computer. Subjects in the control condition did not read anything about the use of a clinical decision aid.

Dependent Measures

After reading the medical scenario, subjects were asked to indicate, using a 5-point rating scale 1) the *quality* of the doctor's decision and 2) whether they would *recommend* this doctor to a friend. Subjects receiving a negative outcome were also asked to indicate 3) how *negligent* they felt the doctor was and 4) how likely they would be to file a *lawsuit* against the doctor. The last question was used as a concrete assessment of perceived responsibility.

Results

A 2-way analysis of variance (outcome × condition) was performed for the 2 measures (decision quality and recommendation) that were used in both the positive and negative outcome conditions. Because the remaining 2 measures (negligence and lawsuit) were recorded only for subjects in the negative outcome condition, they were analyzed via a *t* test for the effects of (computer) condition. Table 1 provides means and standard deviations for all conditions.

Decision Quality

There was no main effect for condition, $F(1, 116) < 1$, $P = 0.69$, and a large main effect of outcome, $F(1, 116) = 49.3$, $P < 0.001$, *partial* $\eta^2 = 0.30$. Decisions were rated more favorably after a positive outcome than after a negative outcome. However, a significant interaction indicated that the effect of outcome on rated decision quality was significantly reduced when subjects were told that the doctor used a decision aid, $F(1, 116) = 8.98$, $P = 0.003$, *partial* $\eta^2 = 0.072$ (Figure 1, top).

Recommend to Friends

There was no main effect for condition, $F(1, 115) = 1.5$, $P = 0.22$, and a large main effect of outcome, $F(1, 115) = 34.3$, $P < 0.001$, *partial* $\eta^2 = 0.23$. Doctors were more likely to be recommended after a positive outcome than after a negative outcome. Again, however, a significant interaction indicated that the effect of outcome information was significantly reduced when sub-

Table 1 Effects of Heeding Recommendation of a Decision Aid on Ratings of Doctor's Decision Quality and on Recommendations of Doctor to Friends (Experiment 1)

Condition		Outcome: Negative	Outcome: Positive
		Decision Quality	
Computer	$\bar{x}$	2.69	3.47
	s	1.07	1.17
Control	$\bar{x}$	2.03	3.97
	s	0.94	1.03
		Recommend to Friends	
Computer	$\bar{x}$	2.07	2.5
	s	0.92	1.08
Control	$\bar{x}$	1.62	3.42
	s	0.78	1.29

Note: Scores range from 1 to 5.

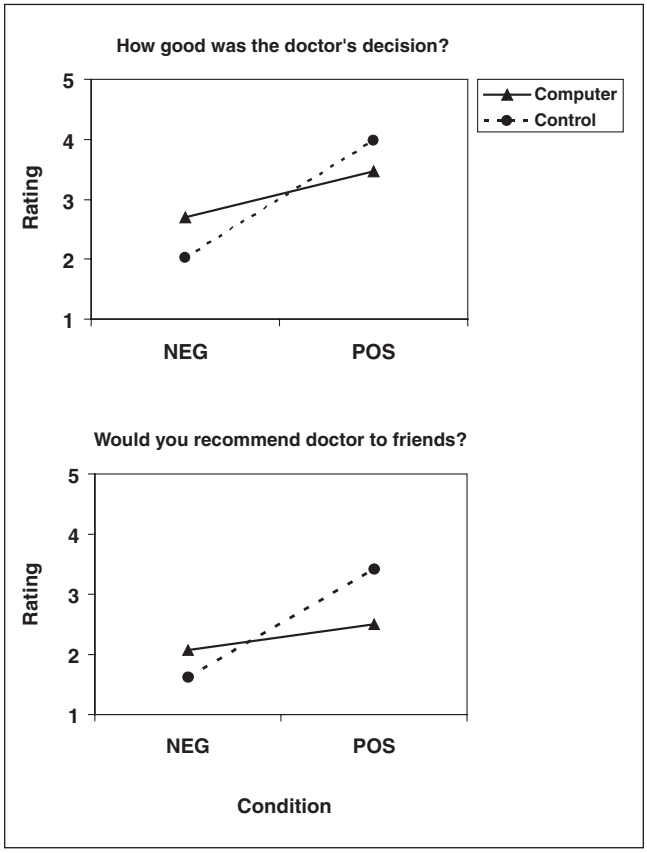

Figure 1 Undergraduates' judgments of doctor's decision quality and likelihood of recommending to others (Experiment 1).

jects were told of the decision aid, $F(1, 115) = 12.92$, $P < 0.001$, *partial* $\eta^2 = 0.10$ (Figure 1, bottom).

Negligence

After a negative outcome, the doctor who followed the advice of a decision aid was judged significantly less negligent ($\bar{x} = 2.86$, $s = 1.16$) than in the control condition ($\bar{x} = 3.48$, $s = 0.99$), $t(56) = 2.20$, $P = 0.032$, $d = 0.58$.

Likelihood of Lawsuit

Subjects were equally likely to believe that they would sue the doctor in both the control and computer decision aid conditions ($\bar{x} = 53.7$, $s = 35.5$), $t(56) < 1$, $P = 0.55$, $d = 0.16$.

Discussion

Use of a computerized decision aid was found to reduce the effects of outcome information on judgments concerning a doctor's competence. As expected, when no mention was made of a computer decision aid, decision quality ratings and likelihood of recommendations to friends were lower after a negative outcome and higher after a positive outcome. These effects were significantly reduced when the doctor used a decision aid. It appears that people may give a decision aid some of the credit for a positive outcome and some of the blame for a negative outcome. After a missed diagnosis, the doctor who had used a computer decision aid was judged as significantly less negligent than one who did not use the decision aid. Such a finding is consistent with the literature on discounting effects.[22] There was no significant difference between subjects in the computer and control conditions in their stated propensity to sue the doctor to recover costs for the surgery. Thus, although use of a decision aid may mitigate some of the negative perceptions of the doctor after a poor outcome, it may not reduce the likelihood of a lawsuit. This finding is somewhat consistent with Kadous's[23] suggestion that jurors consider different factors in determining damage awards than when assessing responsibility.[21]

In Experiment 1, although the doctor considered other diagnoses, he or she always heeded the advice of the decision aid if one was mentioned. However, in practice there are multiple possible interactions with the aid: A physician can agree with the aid, he or she can disagree with the aid but heed it anyway, or the physician can disagree and defy its advice. Thus, Experiment 2 examines each of these possibilities in addition to using no aid at all. In addition, the extent to which physicians are willing to heed or defy a decision

aid may depend on how negatively they believe they will be perceived by laypeople for their decision. Thus, Experiment 2 asked medical students (a reasonable proxy for physicians) to indicate the judgment they expected from patients in addition to their own judgments. Because we focus on perceptions of blame, this experiment only uses negative outcomes.

EXPERIMENT 2

Method

Subjects

First- and 2nd-year medical students ($n = 177$) at the University of South Florida volunteered for this experiment. Of these, 11 failed to complete the questionnaire, leaving 166 subjects. An additional 154 undergraduate students at the University of South Florida St. Petersburg, aged 18 to 66 years ($\bar{x} = 23.5$, $s = 8.5$), received course credit for participation. Data from a total of 320 subjects were analyzed.

Materials and Design

All subjects read a scenario describing a medical malpractice case in which a radiologist was accused of negligence leading to a cancer patient's death. We included variations in the doctor's use of a computerized decision aid to make the diagnosis. In the control condition, no decision aid was mentioned. In the remaining conditions, the doctor either agreed with the aid's diagnosis (agree), initially disagreed with the aid but heeded its diagnosis anyway (heed), or disagreed with the aid and did not heed its diagnosis (defy). Half of the scenarios were cases of omission (the doctor missed a cancer diagnosis), and half were cases of commission (the doctor falsely diagnosed cancer, and unnecessary surgery was performed). Because these 2 conditions were not completely comparable (e.g., differences in cause of death, number of doctors involved in case), we did not analyze for effects of omission/commission.

Dependent Measures

Subjects were asked to indicate on a 7-point scale the following: 1) to what extent the radiologist was at fault for the patient's death, 2) how competent the radiologist was, and 3) how appropriate is it to use computer decision aids in medical decisions. The medical students were asked to make these judgments both as they themselves would and as they thought a typical layperson would. The order of presentation for the 2 types of judgments was counterbalanced across subjects.

Table 2 Effect of Judge and Condition on Perceived Fault and Competence (Experiment 2)

Factor	*F*	*P*	*Partial* η^2
		Fault	
Judge (J)	51.56	<0.001	0.14
Condition (C)	6.62	<0.001	0.06
J × C	<1	0.79	0.01
		Competence	
J	35.65	<0.001	0.10
C	9.13	<0.001	0.08
J × C	<1	0.46	0.00
		Appropriateness	
J	10.76	<0.001	0.03
C	<1	0.48	0.01
J × C	1.87	0.13	0.02

Note: $df_{Error} = 312$, $df_{Judge} = 1$, $df_{Condition} = 3$, $df_{J\times C} = 3$.

Results

First- and 2nd-year medical students did not significantly differ on any of the measures and were combined to form a single medical student group roughly equal in size to the undergraduate group. The resulting design was a 2 (*Judge*: medical student v. undergraduate) × 4 (*Condition*: agree, heed, defy, control). The 3 dependent measures were each subjected to an analysis of variance using this 2 × 4 between-subjects design (Table 2).

Fault

Overall, undergraduates found the doctor to be significantly more at fault ($\bar{x} = 4.69$, $s = 1.65$) than did medical students ($\bar{x} = 3.40$, $s = 1.69$). Most important was a significant main effect of condition. Least squares difference post hoc tests (Table 3) indicated greater fault was perceived when the doctor defied the decision aid than when the doctor agreed with or heeded its advice. Although there was a marginally significant effect of greater fault in the defy condition than in the no-aid control condition ($P = 0.06$, $d = .30$), this effect was not significant at the conventional 0.05 level, even with a very large sample size. Notably, the doctor was judged to be significantly more at fault in the control condition than in the agree condition, which seems to indicate that the use of a computer aid in the context of a medical error is protective if the physician agrees with the aid but not if he or she defies it.

Table 3 Effect of Doctor's Agreement with Decision Aid on Perceived Fault and Competence (Experiment 2)

Measure		Condition			
		Defy (n = 78)	Control (n = 84)	Heed (n = 78)	Agree (n = 80)
Fault	$\bar{x}$	4.62[a]	4.10[a,b]	3.90[b]	3.50[b]
	s	1.73	1.73	1.77	1.77
Competence	$\bar{x}$	4.49[a]	4.33[a]	5.05[b]	5.35[b]
	s	1.55	1.60	1.49	1.26

Note: Row means *not* sharing a superscript are significantly different at 0.05 level.

Competence

Overall, undergraduates found the doctor to be significantly less competent ($\bar{x}$ = 4.31, s = 1.57) than did medical students ($\bar{x}$ = 5.25, s = 1.38). A significant main effect of condition was similar (though not identical) to that found for the fault measure. As can be seen in Table 3, the doctor was perceived to be more competent if he or she either agreed with the computerized decision aid or ultimately heeded its advice than if he or she defied its advice or if no mention of a decision aid was made. No interactions reached significance (Ps > 0.25).

Appropriateness

All subjects indicated that it is moderately appropriate to use a computerized decision aid in general ($\bar{x}$ overall = 5.1, s = 1.39). However, undergraduates felt that it is less appropriate ($\bar{x}$ = 4.84, s = 1.48) than did medical students ($\bar{x}$ = 5.33, s = 1.26). The main effect of condition and the interaction were both nonsignificant.

Medical Students' Predictions of Laypersons' Responses

Medical students had a tendency to find the doctor less at fault and more competent overall than did undergraduates. However, these same medical students were not unaware of the less favorable judgments of a typical layperson. Medical students' predictions for the pattern of lay responses were quite similar to undergraduates' actual responses across the various conditions. Nonsignificant differences between medical students' predictions of and undergraduates' actual ratings were found for fault (P = 0.15, *partial* η^2 = 0.007) and decision aid appropriateness (P = 0.53, *partial* η^2 = 0.001). Medical students did, however, predict that lay judges would rate the doctor's competence significantly lower ($\bar{x}$ = 3.88, s = 1.58) than undergraduates actually did ($\bar{x}$ = 4.32, s = 1.57), $F(2, 301) = 6.26$, P = 0.02, *partial* η^2 = 0.02. The judge × condition interaction was not significant for any of the 3 measures, all Ps > 0.36, all *partial* η^2 < 0.01.

Gender and Age Effects

Gender and age were not recorded for medical students. Undergraduates' ages were similarly distributed across all 4 conditions and were not related to judgments of fault, competence, or appropriateness, all Ps > 0.33, nor did age emerge as a significant covariate when testing the effects of gender and condition on these measures, Ps > 0.13. Female undergraduates rated the doctor significantly more at fault ($\bar{x}$ = 4.9, s = 1.6) than did male undergraduates ($\bar{x}$ = 4.1, s = 1.8), $F(1,146)$ = 9.24, P = 0.003, *partial* η^2 = 0.06, but males and females did not differ in ratings of competence (P = 0.22) or appropriateness (P = .14). Most important, gender did not interact with condition for any of the 3 measures, all Ps > 0.08.

Discussion

Medical Students v. Undergraduates

Some of the largest effects in this experiment were produced by differences between medical students' and undergraduates' ratings. Medical students found the physician less at fault and more competent than did undergraduates, and they rated the use of decision aids more appropriate than did undergraduates. More important, medical students indicated that they understood that the general public is not as confident in decision aids as they are. This finding suggests that perhaps it is not physicians' lack of confidence in decision aids but their awareness of patients' lack of confidence that translates into a reduced use of the aids.

Agree and Heed Conditions

Participants' judgments of fault and competence were most charitable in the heed and agree conditions. This finding is quite reasonable for the agree condition. Although the diagnosis may have been wrong, the doctor was not "alone" in making it—a computerized decision aid also, and independently, came to the same diagnosis. What is surprising is that the heed condition did not differ significantly from the agree condition. This finding is interesting because although the doctor ultimately went with the same incorrect diagnosis, in the heed condition he or she initially made the correct diagnosis. Whether physicians are punished extra harshly for irresolution is an interesting avenue for future research.

Defy and Control Conditions

Judgments of fault and competence were more harsh in the defy and control conditions. These conditions are not as overtly comparable to each other as are the agree and heed conditions. However, their similar ratings suggest that both conditions focus attention solely on the doctor's diagnostic skills. In both conditions, it is only the doctor who makes the incorrect diagnosis. Although arguably rational, the more critical judgments of a doctor who defies the decision aid are also consistent with fears expressed by some doctors hesitant to use such aids. That doctors who defy the aid are judged more harshly after a negative outcome than in any other condition may suggest to some doctors that if a decision aid is used at all, then its advice must *always* be followed to escape added liability.

Our results suggest that a doctor trying to minimize his or her perceived negligence might feel that if using a decision aid, he or she must always heed its advice. This finding is consistent with the fear that "physicians won't think for themselves" expressed by 18% of respondents opposed to computerized decision aids in a study by Grundmeier and Johnson.[11] Similar concerns have also been expressed about the pitfalls of using a decision aid in an auditing setting. "If an auditing firm makes an expert system available to its auditors and an auditor . . . overrides [that] decision aid's recommendation. Will this be viewed as evidence of a lack of due professional care?"[24] It is not surprising then that Kaplan[25] suggested that physicians' negative attitudes toward decision aids can be traced in part to concerns about how using such a device would harm their professional status and interfere with their professional autonomy. On the other hand, decision makers probably *should* heed the advice of diagnostic aids when they disagree because the decision aid is more likely than the doctor to be correct (or at least consistent).[26]

Our findings are generally consistent with recent research concerning auditors' use of decision aids to detect corporate fraud.[21] Using conditions roughly comparable to our defy, control, and heed conditions, they found that auditors who overrode the aid (defy) were found most responsible for the audit failure, followed by auditors with no computerized decision aid (control) and finally auditors who followed the advice of the aid (heed). However, these differences were significant only when subjects believed that the decision aid was correct "about 90 percent of the time."[21(p 193)] When the reliability was lowered to 81%, the difference between these conditions disappeared. In our Experiment 2, when a decision aid was mentioned, subjects were informed that the decision aid was correct about 70% of the time, which may partially explain why there was only a marginally significant difference between our defy and control conditions.

GENERAL DISCUSSION

If it is true that physicians are reluctant to use decision aids because they fear an increased liability in the event of a poor outcome, our research suggests that this fear may not be warranted under certain conditions. In 2 studies, doctors who used and agreed with a computer decision aid were judged less harshly after an error than were doctors who did not consult a decision aid. Experiment 2 showed that the physician may even be protected if the physician heeds the incorrect advice of an aid when his or her personal opinion was correct. The only situation in which a decision aid did not seem to be beneficial was when a doctor defied the correct advice of the aid; however, a physician who defied the aid was perceived no more negatively than the physician who used no aid at all. Thus, it appears that using a decision aid may benefit physicians in the setting of a medical error. The only situation in which the use of decision aid seems to be a clear disadvantage is a positive outcome; in this case, doctors were judged less favorably if an aid was used in the decision.

Our results suggest that a fear of increased perceptions of blame may not be warranted. Furthermore, any fear that does exist may not fully explain physicians' reluctance to use decision aids. Medical students expected laypeople to judge physicians just as laypeople actually did—no more severely with an aid than without one. Because we did not ask medical students how they expected laypeople to respond after a positive outcome, we do not know whether they realize they are judged less positively for a good outcome if a decision aid is used. However, if we assume that medical students are as accurate about laypeople's responses with positive outcomes as they are with negative outcomes, perhaps it is the reduction in favorable judgments that motivates the reluctance to use decision aids. Future research should further examine this issue.

Limitations

Because Experiment 2 examined only negative outcomes, the doctor always made the wrong decision. To manipulate agreement with the decision aid, the aid's diagnosis had to vary across the agreement conditions. Thus, in the heed and agree conditions, the aid made the wrong diagnosis; but in the defy condition, the aid made the correct diagnosis. It is impossible to know the extent to which the difference between the defy condition and other conditions is caused by our manipula-

tion versus the fact that the decision aid made different diagnoses. We believe, however, that this limitation may be a price that researchers asking applied questions must pay to make their designs ecologically valid. It is interesting that the same problem was encountered by Lowe and others[21] who studied decision aid use among auditors. We note optimistically that they provided evidence that judges probably pay more attention to the agreement (or lack thereof) between the decision maker and decision aid than to the actual diagnosis made by the aid.

Experiment 2 used 1st- and 2nd-year medical students to represent the medical community and undergraduates to represent the typical layperson. Although not yet MDs, our medical student sample was certainly more closely aligned with the medical profession than any other non-MD sample we could obtain. In addition, the use of undergraduates may seem an unlikely proxy for the typical medical patient. However, we note that undergraduates at our university are nontraditional in a number of respects. Besides ranging in age from 18 to 66 years ($\bar{x}$ = 24 years), most of our undergraduates work, many have children and/or are married. In addition, the absence of any age effects or interactions with gender suggest that older samples and those that have fewer females are likely to produce similar results. These representations clearly are not perfect, but we consider them reasonable for the goals of the present work.

Finally, because our results are based on only 2 medical cases, it is possible that the findings are particular to those cases and not medical outcomes in general. Although this and other limitations make our conclusions preliminary, they are important because they present a counterexample to conventional wisdom concerning decision aids.

Future Research

We examined the effects of agreement or disagreement with the advice of a decision aid. Future research should also examine the effects of declining to seek the advice of a decision aid. As mentioned earlier, failure to consult an available aid that would have provided important information may still create liability concerns. In addition, further research might attempt to clarify the role of reduced satisfaction after a correct diagnosis in physicians' reluctance to use decision aids.

Finally, if it is true that decision aids deflect some of the responsibility for a poor outcome away from the physician, to whom is that responsibility referred? Future studies could explore whether patients blame the manufacturer of the aid, the hospital board who implemented the aid's use, or perhaps no one at all.

REFERENCES

1. Kohn LT, Conigan JM, Donaldson M. To Err is Human: Building a Safer Health System. Report of the Institute of Medicine. Washington, DC: National Academy Press; 1999.
2. McDonald CJ, Weiner M, Hui SL. Deaths due to medical errors are exaggerated in Institute of Medicine Report. JAMA. 1999;284:93–5.
3. Leape LL. Institute of Medicine medical error figures are not exaggerated. JAMA. 2000;284:95–7.
4. de Dombal FT, Leaper DJ, Staniland JR, McCann AP, Horrocks JC. Computer-aided diagnosis of acute abdominal pain. BMJ. 1972;2:9–13.
5. Bates DW, Leape LL, Cullen DJ, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. JAMA. 1998;280:1311–6.
6. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA. 1998;280:1339–46.
7. Murff HJ, Gandhi TK, Karson AK, et al. Primary care physician attitudes concerning follow-up of abnormal test results and ambulatory decision support systems. Int J Med Inform. 2003;71:137–49.
8. Trivedi MH, Kern JK, Marcee A, et al. Development and implementation of computerized clinical guidelines: barriers and solutions. Methods Inf Med. 2002;41:435–42.
9. Wendt T, Knaup-Gregori P, Winter A. Decision support in medicine: a survey of problems of user acceptance. Stud Health Technol Inform. 2000;77:852–6.
10. Johnston JM, Leung GM, Tin KYK, Ho LM, Lam W, Fielding R. Evaluation of a handheld clinical decision support tool for evidence-based learning and practice in medical undergraduates. Med Educ. 2004;38:628–37.
11. Grundmeier R, Johnson K. Housestaff attitudes toward computer-based clinical decision support. Proceedings of the Annual Meeting of the Medical Informatics Association, Washington, DC. 1999;266–70.
12. Boatsman JR, Moeckel C, Pei BK. The effects of decision consequences on auditors' reliance on decision aids in audit planning. Organ Behav Hum Decis Process. 1997;71:211–47.
13. Ashton RH. Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback and justification. Journal of Accounting Research. 1990;28:148–80.
14. 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.
15. Powell JL. An attempt at increasing decision rule usage in a judgment task. Organ Behav Hum Decis Process. 1991;48:89–99.
16. Bockenholt U, Weber EU. Use of formal methods in medical decision making: a survey and analysis. Med Decis Making. 1992;12:298–306.
17. Sieck WR, Arkes HR. The recalcitrance of overconfidence and its contribution to decision aid neglect. Journal of Behavioral Decision Making. 2005;18:29–53.
18. Meehl PE. Clinical versus statistical prediction: a theoretical analysis and a review of the evidence. Minneapolis: University of Minnesota Press; 1954.
19. Marsh FH. Medical information systems. In: Reich W, ed. Encyclopedia of Bioethics. 2nd ed. New York: MacMillan; 1992.
20. Lowe DJ, Reckers PM. The use of foresight decision aids in auditor's judgments. Behavioral Research in Accounting. 2000;12:97–118.

21. Lowe DJ, Reckers PM, Whitecotton SM. The effects of decision-aid use and reliability on jurors' evaluations of auditor liability. The Accounting Review. 2002;77:185–202.
22. Kelly HH. Attribution in Social Interaction. New York: General Learning Press; 1971.
23. Kadous K. The effects of audit quality and consequence severity on juror evaluations of auditor responsibility for plaintiff losses. The Accounting Review. 2000;75:327–41.
24. Messier WF. Research in and development of audit decision aid. In: Ashton RH, Ashton AH, eds. Judgment and Decision-Making Research in Accounting and Auditing. New York: Cambridge Press; 1995. p 207–28.
25. Kaplan B. Culture counts: how institutional values affect computer use. MD Comput. 2000;17:23–6.
26. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. Science. 1989;243:1668–74.