Journal of Experimental Psychology: Applied
2006, Vol. 12, No. 3, 179–195

Copyright 2006 by the American Psychological Association
1076-898X/06/$12.00 DOI: 10.1037/1076-898X.12.3.179

# Physician Decision Making and Cardiac Risk: Effects of Knowledge, Risk Perception, Risk Tolerance, and Fuzzy Processing

Valerie F. Reyna
Cornell University

Farrell J. Lloyd
Mayo Clinic

Despite training, professionals sometimes make serious errors in risky decision making. The authors investigated judgments and decisions for 9 hypothetical patients at 3 levels of cardiac risk, comparing student and physician groups varying in domain-specific knowledge. Decisions were examined regarding whether they deviated from guidelines, how risk perceptions and risk tolerances determined decisions, and how the latter differed for knowledge groups. More knowledgeable professionals were better at discriminating levels of risk according to external correspondence criteria but committed similar errors in disjunctive probability judgments, violating internal coherence criteria. Also, higher knowledge groups relied on fewer dimensions of information than did lower knowledge groups. Consistent with fuzzy-trace theory, experts achieved better discrimination by processing less information and made sharper all-or-none distinctions among decision categories.

*Keywords:* medical decision making, expertise, risk perception, decision threshold, fuzzy-trace theory

Professionals are often required to make decisions under uncertainty. For example, professors must decide whether graduate students have passed, passed with distinction, or failed qualifying examinations; pilots must decide whether to abort a takeoff, continue on, or take evasive action (e.g., if another plane appears to be in the flight path); and psychiatrists must make decisions about whether patients should discontinue, continue, or change treatments for mental illness (e.g., if a patient is experiencing side effects). In addition to professional training, rubrics or guidelines exist to assist professionals in making such decisions (e.g., Bauer et al., 1999; Braunwald et al., 1994, 2000, 2002; Depression Guideline Panel, 1993; Federal Aviation Administration, 1994). Nevertheless, substantial practice variation remains among professionals, and overall performance falls short of optimality (e.g., Reyna & Adam, 2003; Tierney, Overhage, Takesue, Harris, & Murray, 1995; Wennberg, 1987). For example, 80% of fatal general aviation accidents are a result of pilot error, and medical errors are the eighth leading cause of death in the United States (Kohn, Corrigan, & Donaldson, 2000; U.S. General Accounting Office, 2001). Although prior research has addressed human error (e.g., Fitts, 1947; Reason, 1990), such practice variation, especially for life-and-death medical decisions, is not fully understood. Moreover, decisions involving medical risk allow for examination of broader theoretical claims about expertise, risk perception, and decision making.

Although practice variation—variation in medical or clinical treatment of similar problems—has been well documented, the causes and correlates of that variation are poorly understood (Lee, Pearson, & Johnson, 1995; Weingarten, 1997). We focus on practice variation in identifying levels of cardiac risk for several reasons. A key reason is that once risk is accurately identified, medical and surgical treatments are available that substantially reduce cardiac mortality. In contrast to some diseases, accurate identification of cardiac risk has demonstrable benefits for patients; improved medical outcomes have been demonstrated for treatments ranging from cardiac rehabilitation (using exercise) to medication to surgical intervention, provided that patients are given treatments based on their cardiac risk (e.g., Braunwald et al., 1994, 2000, 2002; Streuber, Amsterdam, & Stebbins, 2006). In addition, heart disease is the major cause of death in the United States, and its impact on society is increasing because of the aging of the population. For example, more than 6 million people per year arrive at emergency rooms complaining of chest pain and discomfort (e.g., Graves & Gillum, 1996). It is estimated that over 20,000 of these patients are sent home erroneously and die of a heart attack (e.g., Graves & Gillum, 1996; see also Collinson, Premachandram, & Hashemi, 2000). Analysts have speculated that many of the remainder are not, on the basis of the latest research on optimal care, *triaged* appropriately (discharged or admitted to the right level of care). Thus, judgment of cardiac risk is a large and important problem, to which psychological research has substantial potential application (Lloyd & Reyna, 2001a; Reyna, 2004).

As with many high-stakes decisions, decision making involving cardiac risk presents certain dilemmas. Admitting everyone with symptoms vaguely suggestive of heart disease to a hospital is not feasible because of the large cost and risk to patient health of inappropriate admission, including unneeded exposure to risks of

Valerie F. Reyna, Departments of Human Development and Psychology, Cornell University; Farrell J. Lloyd, Center for Education Technology, Mayo Clinic Rochester, Minnesota.

The research presented here was partially supported by grants from the National Science Foundation and the Academic Medicine and Managed Care Forum (funded in part by Schering Plough Integrated Therapeutics). We gratefully acknowledge the support of Bruce Jarrell and Raeburn Evans throughout the course of this work.

Correspondence concerning this article should be addressed to Valerie F. Reyna, Department of Human Development, MVR B44, Cornell University, Ithaca, NY 14853. E-mail: vr53@cornell.edu

invasive tests and to antibiotic-resistant hospital-acquired infections. Despite the desirability of discharging low-risk patients (with subsequent follow-up) for chest pain and many other symptoms (e.g., those suggesting stroke or pulmonary embolism), physicians are caught in a quandary as to how admissions can be safely reduced in the face of uncertainty about adverse outcomes. (Again, these considerations apply broadly to any decision that involves uncertainty and the possibility of serious adverse outcomes, such as acting on a potential terrorist threat or releasing a convicted felon into society.) Clinical guidelines have been offered as a practical solution to this quandary (e.g., Braunwald et al., 1994, 2000). Guidelines are not intended to provide hard-and-fast decision rules, but they can reduce clinical uncertainty—for example, by providing clear indications for safely discharging patients. The translation of guidelines into provider behavior is not trivial, however. In fact, psychological research indicates that there are predictable barriers to the successful implementation of clinical guidelines and to improved decision making (e.g., Dawes, 1988; Eddy, 1996; Kahneman, Slovic, & Tversky, 1982; Reyna & Brainerd, 1995; Yates, 1990).

First, some cardiac conditions are more difficult to diagnose than others. Unstable angina is one such condition, and it is a prototype for the kinds of uncertainty that frequently characterize professionals' decisions (e.g., Katz, Griffith, Beshansky, & Selker, 1996). In unstable angina, the patient is not having a heart attack and typically has a normal electrocardiogram. (Because of the subtlety of unstable angina symptoms, cardiologists are commonly consulted; e.g., emergency room physicians may discuss a case with the cardiologist on call before discharging a patient.) However, there are reliable indicators (codified in national guidelines) that place patients at varying levels of imminent risk of having a heart attack (e.g., Braunwald et al., 1994, 2000, 2002). Psychologically, judgments are challenging because indicators are uncertain; medical studies do not pinpoint exact probabilities associated with each indicator (e.g., Dawes, 1988). Thus, ordinal discrimination of risk (i.e., into low-, intermediate-, or high-risk categories) can be achieved, but exact probabilities are elusive.

Results from experiments using an array of tasks suggest that decision makers reason at the least precise level of qualitative gist necessary to accomplish a task, such as the task of deciding among levels of triage (e.g., discharge, admit to a hospital ward bed, or admit to intensive care; for reviews, see Reyna, 2004; Reyna & Brainerd, 1995). Thus, discriminating vague categories of risk so as to map them onto three corresponding levels of triage (i.e., imprecision per se) is a natural and intuitive approach to thinking about risk, one that is taken in the unstable angina guidelines as well as other clinical guidelines aimed at practical issues of diagnosis and treatment (e.g., Reyna, Lloyd, & Brainerd, 2003). According to fuzzy-trace theory, as people become more knowledgeable and experienced at a task, their information processing becomes more gist-based. That is, they increasingly tend to process information as simply, qualitatively, and categorically as possible given the constraints of the task. Although it takes a great deal of expertise to know which details to ignore when evaluating a patient, the prediction is that more expert physicians will base admission decisions on a few, key dimensions of information, such as imminent risk of a heart attack (a myocardial infarction [MI]), rather than the entire panoply of details that less expert physicians will bring to bear on such decisions.

Fuzzy-trace theory also predicts that more knowledgeable decision makers should be able to discriminate better among categories—despite using fewer dimensions of information—than are less knowledgeable decision makers (Reyna & Brainerd, 1995; Reyna et al., 2003). Thus, the most advanced (most experienced and knowledgeable) decision makers should make sharper all-or-none distinctions among categories of risk. As development progresses (e.g., expertise increases), according to the theory, decision making should be based increasingly on vague, intuitive gist categories and less on trading off of exact magnitudes of risk and outcomes (e.g., Adam & Reyna, 2005; Reyna & Ellis, 1994; Reyna & Farley, in press). This developmental prediction has been upheld in a variety of tasks, such as framing tasks and probability judgment, comparing children with adults. Reliance on gist has numerous advantages for human performance, including ensuring that reasoning depends on relatively stable and flexible memory representations (e.g., Reyna & Brainerd, 1991). Nevertheless, it is reasonable to question (and we investigate this question in the current article) whether this developmental prediction holds in comparing adult novices with experts making medical decisions.

In addition to differences in risk perception among decision makers, and thus in discrimination among levels of risk, it is well-known that there are also individual differences in risk tolerance (e.g., Reyna & Farley, in press). For example, Nightingale (1987) found that physicians in a general medical clinic who were loss averse on a risk preference scale ordered twice as many laboratory tests in hypothetical cases as did their colleagues who were not as sensitive to loss. According to Taylor (2000), the more loss averse an emergency department physician is, the more cases are admitted to intensive care units and the longer resuscitation efforts are continued after spontaneous contractions have ceased. Thus, some decision makers have a low tolerance for risk, and such physicians would be expected to be more likely to admit a patient at lower levels of cardiac risk relative to physicians who are not risk averse (see also Swets, Dawes, & Monahan, 2000).

In principle, risk discrimination and risk tolerance are separate influences on decision making and would be expected to each contribute unique variance in, for example, regression equations relating them to decision making (von Winterfeldt & Edwards, 1986; Yates, 1990). The fact that these influences may be separable can be appreciated by imagining that two physicians could perceive identical risk for the same patient, but one (being willing to tolerate that risk) might send the patient home, and the other (being unwilling to tolerate the same level of risk) might admit the patient.

More specifically, we would expect lower tolerance for risk among less knowledgeable decision makers. In other words, less knowledgeable decision makers would be less confident in their risk judgments and, therefore, more likely to hedge by, for example, admitting patients at lower levels of risk. (Research shows that experts report higher confidence overall than do those with less expertise [e.g., Chapman & Chapman, 1969; Shanteau, 1969; Yates, 1990].) Because of wider bands of uncertainty around the judgments of less knowledgeable decision makers at all levels of risk, they would also be more likely to admit patients to intermediate levels of care (neither discharging patients with troubling symptoms nor admitting them to costly intensive care but playing it safe by admitting them to a hospital; von Winterfeldt & Edwards, 1986). Thus, prior research and theory suggests that less knowledgeable decision makers should have lower risk thresholds (i.e., they should admit patients at lower levels of risk, all factors being equal) and should be more likely to hedge by admitting patients to intermediate levels of care. Note that if less knowledgeable decision makers are also less accurate, hedging is entirely appropriate (e.g., Poses et al. 1993; Swets, 1992).

Thus far, we have discussed a number of differences that we would expect to observe between more and less knowledgeable decision makers, including differences in risk discrimination, risk tolerance or threshold, number of decision dimensions processed, and use of hedging (among less knowledgeable decision makers) versus sharply distinguishing categories of patients (among the most expert decision makers). Each of these predictions is developmental in the sense that decision makers with less knowledge and experience, such as medical students, are predicted to differ from those at higher developmental levels, such as cardiologists (e.g., Reyna & Brainerd, 1994).

Some of these developmental differences, such as better discrimination, can be said to reflect correspondence criteria for rationality in that they accurately reflect reality or outcomes in the real world (Adam & Reyna, 2005; Doherty, 2003; Reyna & Adam, 2003). Accurate discrimination among risk categories, for example, can be judged against standards that reflect real differences in risk (e.g., Fischhoff, Lichtenstein, Slovic, Derby, & Keeney, 1981; Johnson, 1993; Loewenstein & Mather, 1993; Slovic, Malmfors, Krewski, & Mertz, 1995; Smith & Kida, 1991). Correspondence criteria for rationality have been distinguished from coherence criteria that do not refer to external reality but, rather, to internal coherence or consistency among an individual's judgments and/or decisions. According to fuzzy-trace theory, correspondence and coherence criteria are fundamentally different from each other (though both are essential for a full account of rationality); the former are expected to differ developmentally (as already discussed), but specific types of coherence difficulties are not predicted to differ developmentally.

The reason for the lack of developmental differences in internal coherence is that many illusions or biases that violate coherence are produced by processing interference between overlapping classes (e.g., Reyna, 1991; Reyna & Brainerd, 1993; Sloman, Over, Slovak, & Stibel, 2003). Processing interference errors are the last class of errors to occur in development, and extensive experimentation has shown that they are advanced errors that do not reflect logical, memorial, or conceptual deficits (see Reyna, 1995; Reyna & Brainerd, 1995). Instead, they have to do with mental bookkeeping difficulties in keeping track of overlapping referent classes, compounded by distractions from compelling gists of alternative categories (e.g., patients who have coronary artery disease but are not at risk of an MI or patients who have coronary artery disease but are at risk, each of which fit a compelling prototype for clinicians, in contrast to patients at risk of an MI without coronary artery disease [an atypical presentation]). Indeed, research has shown that susceptibility to compelling gists often increases with development as a result of increased emphasis on meaning as opposed to verbatim or literal information processing, as predicted by fuzzy-trace theory (e.g., Klaczynski, 2005; Reyna & Ellis, 1994). Therefore, in contrast to predictions for correspondence criteria, such as discrimination, coherence errors resulting from processing interference (e.g., from overlapping classes) are predicted to remain robust in more advanced phases of development. In the current study, we sought to determine whether this prediction would hold for highly advanced decision makers, physicians, reasoning in their domain of expertise.

Specifically, we studied variations in domain-specific expertise in unstable angina by comparing judgments and decisions of medical students, family practice physicians, emergency medicine physicians, internal medicine physicians, cardiologists, and nationally recognized expert cardiologists. Each of these groups had received medical training in judging risks and making admission decisions for patients with symptoms suggestive of unstable angina (see Table 1 for examples of symptoms). Naturally, all of the physicians were once medical students, and students have the lowest level of training. Family practice, emergency medicine, and internal medicine physicians are all generalists who are trained to recognize unstable angina (e.g., Braunwald et al., 1994; Harrold, Field, & Gurwitz, 1999; Salerno, Alguire, & Waxman, 2003). However, family practice physicians generally see a younger patient population (i.e., families with young children) and are somewhat less knowledgeable about chronic diseases of aging such as cardiovascular disease. Emergency medicine and internal medicine physicians, however, see cardiovascular disease more routinely as part of their practices (and receive more specialized training in this area). Emergency medicine physicians treat patients who present with chest pains and other symptoms suggestive of heart disease for acute episodes (although they also treat physical trauma and many other problems). Internal medicine physicians are also highly familiar with cardiovascular disease and tend to treat older patients. Cardiology is a subspecialty of internal medicine that requires training beyond that for internal medicine. Hence, the cardiologists in this study were all once internal medicine physicians who received additional training in cardiology. Studies comparing generalist physicians (e.g., family practice physicians) with specialists (e.g., cardiologists) in the domain of specialty generally report differences in knowledge as well as in adherence to practice guidelines and medical outcomes (e.g., Chin, Friedman, Cassel, & Lang, 1997; Harrold et al., 1999; Salerno et al., 2003). (The physician groups were about equally well versed in statistics in that a single course in that area is generally taken in medical school prior to specialization and board certification.) Therefore, we hypothesized that medical students, family practice physicians, emergency medicine physicians, internal medicine physicians, cardiologists, and expert cardiologists would differ on measures related to knowledge of cardiovascular disease, including perceptions of MI risk, perceptions of coronary artery disease (CAD) probability, admission probability, triage decisions, and risk tolerance (the level of risk physicians are willing to tolerate in discharging patients) for patients varying in level of risk according to the unstable angina guidelines (Braunwald, 1994, 2000, 2002).

According to the guidelines, the two key dimensions of risk and probability in diagnosing unstable angina are the probability of clinically significant CAD and the imminent risk of MI. A disjunctive combination of these two judgments should govern admission decisions (Braunwald et al., 1994, 2000): If either the probability of CAD is high or the risk of MI is high, the patient should be admitted to the highest level of care—intensive care. If one or the other is intermediate, the patient should be admitted, but not necessarily to the highest level of care. Low risk for both judgments should result in a decision to discharge the patient (with follow-up by a physician in 72 hr). Thus, perceptions of risk and

Table 1
*Characteristics of the Case Presentations*

| Characteristic | Low-risk cases 1 | Low-risk cases 2 | Low-risk cases 3 | Intermediate-risk cases 4 | Intermediate-risk cases 5 | Intermediate-risk cases 6 | High-risk cases 7 | High-risk cases 8 | High-risk cases 9 |
|---|---|---|---|---|---|---|---|---|---|
| Age (years) | 48 | 35 | 54 | 48 | 40 | 65 | 40 | 54 | 65 |
| Sex | M | M | M | M | F | F+ | F | M | F+ |
| Risk factors | | | | | | | | | |
| Diabetes | No | No | No | No | No | No | No | No | No |
| Hypertension | No | No | Yes | No | No | No | Yes | Yes | Yes |
| Cholesterol | No | No | No | No | No | No | No | No | No |
| Smoking | 29 years | 10 years | No | 20 years | 10 years | 10 years | Past use | Past use | Past use |
| History of CAD | No | No | No | No | No | No | No | No | No |
| Definite angina | No | No | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Pain quality | Sharp; jabbing | Pressure | Sharp; jabbing | Pressure | Pressure | Pressure | Pressure | Pressure | Pressure |
| Frequency | 4–5 in 2 mo. | 4–5 in 2 mo. | 2 in 2 days | 4–5 in 2 mo. | 4–5 in 2 mo. | 4–5 in 2 mo. | 2 in 2 days | 2 in 2 days | 2 in 2 days |
| Pain on exertion | No | Yes | No | Yes | Yes | Yes | No | No | No |
| Pain free at evaluation | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Normal ECG | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Evidence of congestive heart failure | No | No | No | No | No | No | Yes | Yes | Yes |
| Likelihood of CAD | Low | Low | Low | Intermediate | Intermediate | Intermediate | High | High | High |
| Risk of MI or death | Low | Low | Low | Intermediate | Intermediate | Intermediate | High | High | High |
| Overall risk | Low | Low | Low | Intermediate | Intermediate | Intermediate | High | High | High |

*Note.* The cases were not presented to participants in tabular form but as narrative summaries. They were designed to contain key historical data that would place patients in three different triage categories on the basis of the likelihood of coronary artery disease (CAD) and the risk of myocardial infarction (MI) or death as described by the guidelines. Cases were designed so that the triage decision would be based on history and physical exam data only. So that there would be a complete clinical summary, the laboratory work, electrocardiogram (ECG), and chest x-ray were reported as normal. Participants were told to assume that the patient was pain free at the time of the triage decision. M = male; F = female; + = postmenopausal.

probability along these two dimensions ought to, according to guidelines, predict admission decisions (although, as noted, fuzzy-trace theory predicts that the most expert will make decisions using fewer dimensions of information with starker separations among categories, reflecting more simplistic, all-or-none thinking).

However, disjunctive judgments (e.g., of CAD *or* MI) are prone to error because of processing interference, as discussed above, because they involve overlapping classes. People tend to underestimate disjunctive probabilities (and to overestimate conjunctive probabilities; e.g., Tversky & Koehler, 1994). Because the combination of CAD and MI probabilities is subject to systematic biases (Reyna, 1991; Reyna & Brainerd, 1993, 1994, 1995), the process of combining risk estimates is another source of predictable variation in physicians' decisions (Eddy, 1982; Reyna et al., 2003). On the basis of fuzzy-trace theory, we would anticipate that these errors would not vary with knowledge of cardiovascular disease, and such knowledge should not prevent advanced decision makers from making coherence errors.

To briefly review, in this study, we assessed practice variation among physicians for judgments and decisions involving cardiac risk. We did this by comparing physicians' decisions for hypothetical patients who varied in risk, according to national guidelines, with decisions recommended in the guidelines. As a check on our interpretation of the guidelines, we also compared physicians' decisions with those of the nationally recognized expert cardiologists. Our hypothesis was that physicians' decisions would differ significantly from those recommended in the guidelines or by experts, exemplifying practice variation. In analyzing the data, we drilled down to the dimensions of risk and probability that are supposed to underlie physicians' decisions, according to the guidelines, examining effects of knowledge on risk perception (and on the ability to discriminate a patient's level of risk) and whether variations in risk perception and risk tolerance accounted for variation in decision making. We used analyses of variance to analyze effects of knowledge on perceptions of MI risk, perceptions of CAD probability, admission probability, triage decisions, and risk tolerance (the level of risk physicians are willing to tolerate in discharging patients) for patients varying in level of risk. We hypothesized that physician groups would differ in risk tolerance (e.g., because of hedging), in their ability to discriminate between low and high risk of MI and of CAD, and in triage decisions for low- and high-risk patients (i.e., that more knowledgeable physicians would discriminate better between low- and high-risk patients). To explain variation in decisions, we used multiple regression to relate risk perception for MI, perception of probability of CAD, risk tolerance for MI, and risk tolerance for CAD (as well as perception of "other" risks) to predict triage decisions (i.e., level of triage in terms of discharge, admission to the hospital but not to intensive care, or admission to intensive care). We hypothesized that these dimensions of risk perception and risk tolerance should each contribute uniquely to triage decisions. Finally, we also examined such cognitive processes as the number of dimensions considered in decision making (correlating perceived risks with level of triage) and the internal coherence among physicians' disjunctive judgments of risk and probability (analyzing the number of logical errors by level of risk and physician knowledge using analysis of variance [ANOVA]). We hypothesized that more expert decision makers would use fewer dimensions of information, but expertise would not affect the number of logical errors, which were not expected to differ significantly for physician groups.

## Method

### Participants

Following institutional review board approval and informed consent, attending and resident physicians affiliated with an academic health sciences center in the southwest United States, including those in community practice settings, were enrolled in the study. A total of 75 participants (66 physicians and 9 medical students) enrolled in the study. One physician refused to participate. The overall mean age of the participants was 35.9 years ($SD = 9.6$, range = 24–64); 25% of the participants were female. The mean age of the physician sample was 37.0 years ($SD = 9.5$), 27% were female, and their mean self-reported level of experience with unstable angina (on a 1–7 scale) was 4.9 (*highly experienced*; $SD = 1.2$). The mean age of the medical student sample was 27.2 years ($SD = 3.9$), and 22% were female. Physicians were drawn from several areas of specialty pertinent to the diagnosis and triage of unstable angina: family practice ($n = 12$; mean age = 30.8 years, $SD = 4.1$), emergency medicine ($n = 20$; mean age = 33.7 years, $SD = 6.1$), internal medicine ($n = 21$; mean age = 35.6 years, $SD = 7.2$), and cardiology ($n = 13$; mean age = 50.2 years, $SD = 9.0$). Three national experts in cardiology, all of whom were male (mean age = 60.0 years, $SD = 5.3$), responded to the same stimuli using the same procedures to provide an additional benchmark for evaluating performance.

### Procedure and Materials

Participants responded to nine patient descriptions, three at each of three levels of overall risk—low, intermediate, or high—according to guidelines for unstable angina (see Table 1 and the Appendix). The descriptions were fashioned by a practicing internal medicine physician (who did not participate in the study) to capture the information and format typically used by physicians to communicate about cases—for example, when emergency room physicians consult with cardiologists about patients with symptoms suggestive of unstable angina. All patients had normal resting electrocardiograms, laboratory work, and x-rays. The descriptions were presented in random order, and for each patient, the participant made five judgments of risk or probability and a triage decision. The judgments were estimates of (a) the probability that the patient had clinically significant CAD, (b) the patient's imminent risk of MI, (c) the probability of all other conditions that would warrant admission, (d) the overall probability that the patient should be admitted, and (e) the probability that the patient had clinically significant CAD or was at imminent risk of MI. The response scale for judgments ranged from 0 to 100; instructions indicated that "0 means no chance at all, 50 means as likely as not, and 100 means absolutely certain." The options provided for the triage decision were (a) outpatient with follow-up (i.e., discharge the patient from the emergency department), (b) admission to an unmonitored ward bed in the hospital, (c) admission to a telemetry unit (i.e., monitored ward bed) in the hospital, (d) admission to the cardiovascular intensive care unit, or (e) other (specified by the participant).

After responding to each of the nine patient descriptions, participants answered a series of questions about their background (e.g., demographics, area of specialty, level of clinical experience with unstable angina) and attitudes. The response scale for rating clinical experience was anchored by *almost none* (1) and *expert* (7), with *very low* (2), *low* (3), *average* (4), *high* (5), and *very high* (6) as intermediate ratings. Participants were also asked to indicate which of two possible errors in admission decisions is worse—not admitting a patient who really should have been admitted or admitting a patient who really should not have been admitted. They then rated the degree of difference between these two errors on a 0–100 scale (for which 0 indicated *no difference at all* between the two errors and 100 indicated the *maximum possible difference* between the two errors).

## Results

The two intermediate response categories, both referring to hospital admission, were combined for analyses (guidelines do not distinguish between them). "Other" was rarely selected, and such responses (e.g., participants provided a specific hospital unit location) were easily classified into one of the three levels of triage that the guidelines distinguish: 1 = *discharge for outpatient follow-up*, 2 = *admission to the hospital but not to intensive care*, or 3 = *admission to cardiac intensive care*. Initial analyses compared residents and attending physicians, but these groups did not differ significantly, and subsequent analyses collapsed across this factor. Effect sizes were estimated by Cohen's $f$ (Cohen, 1988), except where otherwise noted. Conventionally, an $f$ value of 0.10 is indicative of a small effect, 0.25 is indicative of a medium effect, and 0.40 is indicative of a large effect.

### Overview

In the following sections, we report results of analyses (including goodness-of-fit analyses) regarding whether trained professionals deviated from practice guidelines and from top experts' recommendations. We scored decisions as correct or not and analyzed these scores using ANOVA to examine whether deviations differed according to knowledge group or patient characteristics. Given that decisions deviated, we then explored factors that were hypothesized to determine those decisions—that is, risk perceptions and risk tolerances. We used multiple regression to predict decisions from risk perceptions and risk tolerances, investigating the hypothesis that each of these factors would contribute unique variance to decision making. We then conducted a series of ANOVAs investigating effects of knowledge and patient risk level on risk perceptions, risk tolerances, and decision making. Regarding these ANOVAs, we hypothesized that more knowledgeable professionals would be better able to discriminate between low- and high-risk patients, as reflected in admission decisions and risk perceptions for MI and CAD, and that physicians differing in knowledge would also differ in risk tolerance such that lower knowledge would be associated with lower risk tolerance and with hedging (choosing the middle category) in admission decisions. Each of these analyses involved external criteria, such as deviation from external guidelines and discrimination between patients at different risk levels, according to those guidelines.

We performed additional analyses concerning internal coherence and cognitive processing, including examination of correlations between judgment dimensions and decisions. We hypothesized that fewer dimensions would correlate with decisions for professionals at higher levels of knowledge (i.e., they would process fewer dimensions of information to make decisions). We also examined internal coherence by using ANOVA to analyze the number of coherence errors by knowledge group and patient risk. We hypothesized that even advanced decision makers would be subject to errors of coherence in disjunctive probability judgments because these errors are a result of processing confusions rather than knowledge limitations. Finally, we were able to secure the judgments and decisions of three top experts in cardiology for the nine patients. We used ANOVA to compare the larger sample's decisions with those recommended by the top experts (scoring *correct* and *incorrect* according to the experts' recommendations), again to investigate practice variation. We examined experts' decisions to test our hypothesis that the most advanced decision makers would make sharper all-or-none distinctions among decision categories.

### *Deviations From Clinical Guidelines*

The first question was whether decisions deviated from those recommended in the guidelines. Recall that we hypothesized that there would be significant deviation from guidelines. Goodness-of-fit tests indicated lack of fit between observed decisions and those stipulated by the guidelines for unstable angina. The deviation of observed decisions from the guidelines was significant for each patient, $t(74) = 7.05, 9.70, 18.05, 4.48, 8.05, 2.98, 13.00, 8.51$, and 10.54, all $ps < .01$, for Patients 1–9, respectively. The respective effects sizes, computed as Cohen's $d$, were 1.64, 2.25, 4.20, 1.04, 1.87, 0.69, 3.02, 1.98, and 2.45. (Lack of fit remained when students were excluded from the analyses, and the same pattern of significance was obtained for ordinal tests [i.e., Kolmogorov–Smirnov tests].) The overall percentage of decisions that strictly agreed with the guidelines was 46%. The percentage increased to 67% if decisions were coded simply on the basis of admission versus nonadmission (i.e., if any decision to admit, either to a hospital ward bed or intensive care, for intermediate- and high-risk patients was counted as agreement). (Similar figures were obtained when national experts were used as the standard rather than the guidelines; [see below].) Therefore, our hypothesis of significant deviation was upheld.

We then scored decisions as *correct* (1) or *incorrect* (0) according to the guidelines and input these data into ANOVAs (e.g., Seeger & Gabrielsson, 1968) with knowledge group (five levels: medical students, family practice physicians, emergency medicine physicians, internal medicine physicians, and cardiologists), patient risk (three levels: low, intermediate, and high), and replications within each risk level (three replications) as factors. For both ways of scoring agreement with the guidelines—strict agreement or admission agreement—ANOVAs revealed that agreement differed for patients at different levels of risk (see Tables 2 and 3). Proportion of strict agreement was greatest for the intermediate-risk patients and declined for both low- and high-risk patients (see Table 2). If only admission agreement is considered, however, deviations from the guidelines were asymmetrical: There were more low-risk patients admitted than there were high-risk patients discharged (see Table 2). Indeed, physician agreement with the guidelines was nearly perfect (.98) for the high-risk patients. A main effect of knowledge group was also obtained for strict agreement (see Table 3). Level of patient risk interacted with knowledge group for strict agreement and also for admission agreement (see Table 3). Differences among knowledge groups, and associated interactions, are described in greater detail below where we report group differences in decisions, risk perceptions, and risk tolerances. As also described in greater detail below, patients within risk levels (i.e., replications) were not treated as homogeneous. Replication and the interaction of replication with risk were significant for both strict agreement and admission agreement analyses (see Table 3). Thus, agreement with the guidelines was greatest for high-risk patients (consistent with low risk tolerance, examined in detail below) and was associated with differences in knowledge (also examined below).

Table 2
*Proportions of Strict Agreement and Admission Agreement With Guidelines for Level of Triage for Patients at Low, Intermediate, and High Levels of Risk*

| Patient risk | Strict agreement | Admission agreement |
|---|---|---|
| Low | .38 | .38 |
| Intermediate | .68 | .69 |
| High | .33 | .95 |

*Note.* Strict agreement was based on the guidelines for unstable angina. Admission agreement was based on assessment of either admission decision (ward bed or cardiac intensive care unit) as agreement for intermediate- and high-risk patients.

### *Predicting Decisions from Risk Perceptions and Risk Tolerances*

Given that participants' decisions differed significantly from those recommended in the guidelines, it made sense to investigate the bases for those decisions. We examined how perceptions of MI risk, CAD probability, other risks warranting admission, tolerance of risk of an MI, and tolerance of risk of CAD predicted triage decisions at each level of patient risk. Perceptions of MI risk and CAD probability were obtained from participants' estimates of those quantities, averaged separately for low-, intermediate-, and high-risk patients. For each participant, overall risk tolerance was calculated, separately for MI risk and CAD probability, by taking the highest estimate of MI risk that was coupled with a discharge decision as a measure of MI risk tolerance and the highest estimate of CAD probability coupled with a discharge decision as a measure of CAD risk tolerance. Because the nine patients varied across a wide range of risks and probabilities, we were able to obtain reasonably sensitive measures of risk tolerance, as demonstrated by the detection of significant effects (see below). Table 4 shows the results of multiple regression analyses for low-risk, intermediate-risk, and high-risk patients.

For low-risk patients, all predictors except the estimate of other risks were significant, indicating that estimates of MI risk and CAD probability each contributed to triage decisions (per the guidelines). In addition, tolerance for risk also contributed significantly to decisions, beyond perceptions of risk, and risk tolerances for MI and for CAD contributed independently to admission decisions. As might be expected, beta weights for risk tolerance were negative: The lower the tolerance for risk, the more likely it was that low-risk patients would be triaged to higher levels of care (i.e., admitted to the hospital).

Results for patients at intermediate levels of risk were similar to those for patients at low risk, except that perception of CAD probability failed to reach significance as a predictor of triage decisions. Perceptions of MI risk and risk tolerances for MI and for CAD each contributed unique variance in predicting admission (triage) decisions for patients at intermediate levels of risk. For patients at high levels of risk, perceptions of risk of MI as well as probability of CAD predicted triage decisions, and risk tolerance for MI was also significant. However, risk tolerance for CAD was not a significant predictor of admission decisions for high-risk patients. Patients at high levels of risk exceeded the threshold for

Table 3
*Analysis of Variance Results for Effects of Knowledge Group, Patient Risk, and Replication on Strict Agreement and Admission Agreement With the Guidelines*

| | | Strict agreement | | | Admission agreement | | |
|---|---|---|---|---|---|---|---|
| Source | *df* | *MS* | *F* | Cohen's *f* | *MS* | *F* | Cohen's *f* |
| Group (G) | 4 | 0.64 | 2.62* | 0.39 | 0.14 | 1.16 | 0.26 |
| Error: G | 70 | 0.24 | | | 0.12 | | |
| Risk (R) | 2 | 7.18 | 18.54** | 0.51 | 16.39 | 63.15** | 0.95 |
| G × R | 8 | 1.70 | 4.40** | 0.50 | 1.12 | 4.32** | 0.50 |
| Error: R | 140 | 0.39 | | | 0.26 | | |
| Replication (r) | 2 | 0.80 | 7.61** | 0.33 | 0.63 | 7.76** | 0.33 |
| G × r | 8 | 0.12 | 1.14 | 0.25 | 0.18 | 2.25* | 0.50 |
| Error: r | 140 | 0.10 | | | 0.08 | | |
| R × r | 4 | 2.35 | 19.26** | 0.52 | 2.30 | 24.15** | 0.59 |
| G × R × r | 16 | 0.21 | 1.75* | 0.32 | 0.16 | 1.68 | 0.31 |
| Error: R × r | 280 | 0.12 | | | 0.09 | | |

* $p < .05$. ** $p < .01$.

admission across a range of risk tolerances for CAD (and, so, CAD risk tolerance was not a significant predictor), but the more serious risk, that of MI, remained significant in predicting levels of triage (e.g., in determining whether high-risk patients would be admitted to the highest level of care, cardiac intensive care). Overall, therefore, the hypothesis was confirmed that differences in risk tolerance would predict unique variance in decision making, beyond that accounted for by risk perceptions.

### *Effects of Knowledge on Risk Perceptions, Risk Tolerances, and Decision Making*

Previous analyses indicated that decisions deviated significantly from those recommended in clinical guidelines. This finding is an illustration of practice variation—namely, treating the same or similar patients differently (in this instance, differently from the way the experts who wrote the guidelines would treat these patients). In this section, we present analyses of differences in decision making among knowledge groups designed to address the issue of practice variation. These results are followed by results of analyses of differences among knowledge groups in risk perceptions and risk tolerances that underlie decision making (according to the findings of regression analyses reported above). Each of the analyses presented here involved comparison of knowledge groups (medical students, family practice physicians, emergency medicine physicians, internal medicine physicians, and cardiologists) for patients at three levels of risk (low, intermediate, and high), with three replications within each level of patient risk. Self-reported ratings of experience with unstable angina on a 1–7 scale differed for knowledge groups: Mean ratings were 2.2 for medical students ($SD = 1.1$), 3.8 for family practice physicians ($SD = 0.4$), 4.8 for emergency medicine physicians ($SD = 1.2$), 4.6 for internal medicine physicians ($SD = 0.7$), and 6.6 for cardiologists ($SD = 0.6$).

*Differences in decision making across knowledge groups.* An ANOVA comparing triage decisions for knowledge groups across levels of patient risk and replication showed that knowledge group, patient risk, and their interaction had significant effects on decisions (see Table 5). Patients classified as high risk by the guidelines were more likely to be triaged at higher levels than were patients at low risk (although intermediate- and low-risk patients did not differ significantly). Differences among knowledge groups displayed what developmental theorists often refer to, qualitatively, as a U-shaped curve, in which participants at the developmental extremes display superficially similar behavior. Mean triage levels were lowest for medical students and cardiologists, and they peaked for emergency medicine physicians. However, this main effect of knowledge group was qualified by an interaction

Table 4
*Multiple Regression Analyses for Variables Predicting Level of Triage Decision for Low-, Intermediate- and High-Risk Patients*

| | Low risk | | | Intermediate risk | | | High risk | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | *B* | *SE B* | β | *B* | *SE B* | β | *B* | *SE B* | β |
| CAD probability | 0.01 | 0.00 | .50** | 0.00 | 0.00 | .12 | 0.01 | 0.00 | .31* |
| MI risk | 0.01 | 0.00 | .47** | 0.01 | 0.00 | .58** | 0.01 | 0.00 | .37* |
| Other risks | 0.00 | 0.00 | .04 | 0.00 | 0.00 | −.03 | 0.00 | 0.00 | .02 |
| CAD risk tolerance | 0.00 | 0.00 | −.32** | −0.01 | 0.00 | −.51** | 0.00 | 0.00 | .06 |
| MI risk tolerance | −0.01 | 0.00 | −.49** | −0.01 | 0.00 | −.49** | −0.01 | 0.00 | −.28* |
| $R^2$ | | .55 | | | .58 | | | .31 | |
| *F* | | 16.81** | | | 18.94** | | | 6.25** | |

*Note.* $N = 75$. CAD = coronary artery disease; MI = myocardial infarction (heart attack).
* $p < .05$. ** $p < .01$.

Table 5
*Analysis of Variance Results for Effects of Knowledge Group, Patient Risk, and Replication on Admission Decision and Admission Probability*

| Source | *df* | Admission decision | | | Admission probability | | |
|---|---|---|---|---|---|---|---|
| | | *MS* | *F* | Cohen's *f* | *MS* | *F* | Cohen's *f* |
| Group (G) | 4 | 2.21 | 3.84** | 0.47 | 5,174.51 | 1.19 | 0.26 |
| Error: G | 70 | 0.58 | | | 4,366.98 | | |
| Risk (R) | 2 | 23.43 | 72.72** | 1.02 | 40,265.52 | 54.91** | 0.89 |
| G × R | 8 | 1.10 | 3.42** | 0.44 | 2,561.06 | 3.49** | 0.45 |
| Error: R | 140 | 0.32 | | | 733.37 | | |
| Replication (r) | 2 | 4.46 | 29.98** | 0.65 | 3,459.35 | 7.48** | 0.33 |
| G × r | 8 | 0.23 | 1.54 | 0.30 | 469.51 | 1.02 | 0.24 |
| Error: r | 140 | 0.15 | | | 462.16 | | |
| R × r | 4 | 2.47 | 16.99** | 0.49 | 9,278.92 | 28.96** | 0.64 |
| G × R × r | 16 | 0.23 | 1.56 | 0.30 | 398.87 | 1.24 | 0.27 |
| Error: R × r | 280 | 0.15 | | | 320.41 | | |

** $p < .01$.

with patient risk (see Table 5). As can be seen in Figure 1, the *difference* between triage levels for low- and high-risk patients increased from .33 for both medical students and family practice physicians (at lower levels of domain-specific knowledge) to .78, .80, and .82 for physicians in emergency medicine, internal medicine, and cardiology, respectively (at higher levels of domain-specific knowledge). Replication and the Risk × Replication interaction were significant, but these effects did not interact with knowledge group (see Table 5).

Similar effects of patient risk and the interaction of risk with knowledge group were obtained for judgments of admission probability, although knowledge group was not significant as a main effect (see Table 5). Patients at high risk according to the guidelines were judged as more likely to need admission to the hospital than were those at low risk, with intermediate-risk patients somewhere in the middle (see Table 5). Again, the main effect was qualified by an interaction between patient risk and knowledge group (see Table 5). Although low- and high-risk patients differed for each of the knowledge groups, the interaction was significant because this difference was not identical across knowledge groups. As can be seen in Figure 1, the difference in admission probability between low-and high-risk patients was larger with increasing knowledge: Mean admission probabilities for low- and high-risk patients differed by 10 points for medical students, 23 points for family practice physicians, 30 points for emergency medicine physicians, 33 points for internal medicine physicians, and 40 points for cardiologists. Replication and the Risk × Replication interaction were significant, but again, these effects did not interact with knowledge group (see Table 5). Therefore, the analyses for both admission (triage) decisions and admission probabilities confirmed the hypothesis that knowledge groups would differ in terms of discriminating low- from high-risk patients.

*Differences in risk perceptions across knowledge groups.* The analyses on level of triage decision and on admission probability consistently indicated that participants discriminated among risk categories and that discrimination was greater at higher levels of knowledge. We now turn to analyses that further unpack differences in decision making across knowledge groups. According to the regression analyses presented above, risk perceptions (of MI risk and CAD probability) and risk tolerances (for MI risk and CAD probability) predicted decision making. Thus, given that knowledge groups differed in decision making, we investigated whether these differences could be traced to specific differences in risk perceptions and risk tolerances.

For perceptions of both MI risk and CAD probability, the factor of patient risk (according to the guidelines) was significant (see Table 6). Participants discriminated among levels of patient risk as stipulated by the guidelines in judging MI risk and CAD probability. For MI risk, means for low, intermediate, and high risk were 31% (*SE* = 2.25), 37% (*SE* = 2.70), and 51% (*SE* = 2.82), respectively; for CAD probability, means were 57% (*SE* = 1.97), 68% (*SE* = 1.74), and 72% (*SE* = 1.94), respectively. However, the pattern across knowledge groups for risk perceptions was found to differ. For perceptions of MI risk, there was a main effect of knowledge group but no interaction with patient risk (see Table 6). Cardiologists perceived the imminent risk of MI for these patients (regardless of risk level) to be lower than did other knowledge groups. The mean estimate for cardiologists was 23% (*SE* = 5.36), as opposed to 49% (*SE* = 6.45) for medical students, 36% (*SE* = 5.58) for family practice physicians, 48% (*SE* = 4.32) for emergency medicine physicians, and 42% (*SE* = 4.22) for internal medicine physicians. The opposite pattern was obtained for CAD probability. There was no significant main effect of knowledge group, but there was an interaction between knowledge group and patient risk (see Table 6). Although knowledge groups did not differ overall, differences between low- and high-risk patients in perceived CAD probability were larger for higher knowledge groups: 3 points for medical students, 9 points for family practice physicians, 20 points for emergency room physicians, 19 points for internal medicine physicians, and 23 points for cardiologists. Replication and the Risk × Replication interaction were significant for both MI risk and CAD probability, but again, these effects did not interact with knowledge group (see Table 6). Thus, complementing the earlier decision analyses, risk perceptions differed across knowledge groups such that cardiologists perceived a lower risk of MI in general, and the ability to discriminate between low- and high-risk patients for CAD differed across knowledge groups.

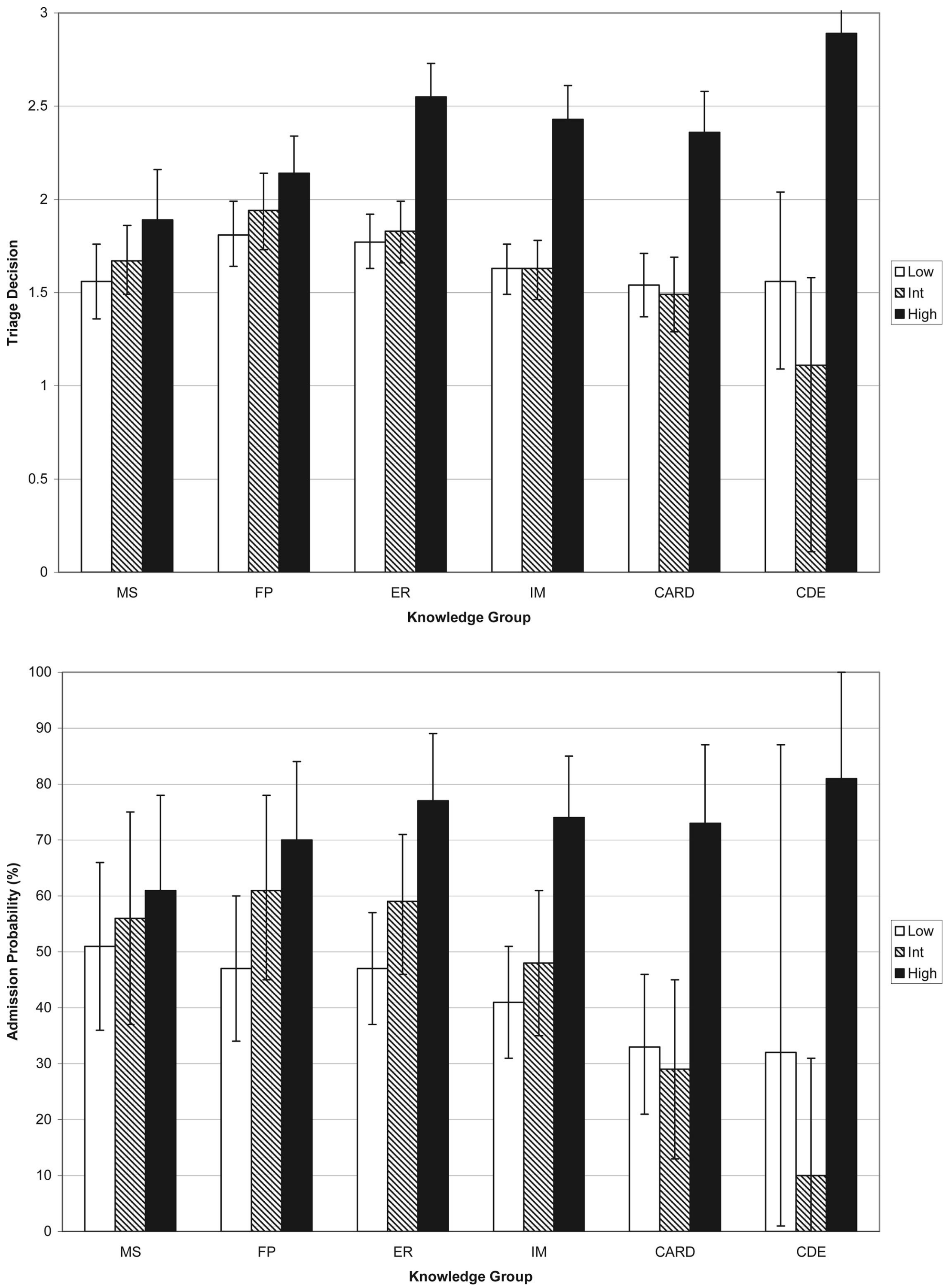


*Figure 1.* Mean levels of triage/admission decision (upper panel; 1 = outpatient, 2 = ward bed, 3 = cardiac intensive care) and admission probability (lower panel; judged overall probability that the patient's condition warrants admission) for patients at low, intermediate (Int), and high levels of risk, presented separately for six groups differing in level of knowledge (MS = medical students, FP = family practice physicians, ER = emergency medicine physicians, IM = internal medicine physicians, CARD = cardiologists, CDE = cardiology experts). Error bars represent 95% confidence intervals.

Table 6
*Analysis of Variance Results for Effects of Knowledge Group, Patient Risk, and Replication on Estimates of MI Risk and CAD Probability*

| | | MI risk | | | CAD probability | | |
|---|---|---|---|---|---|---|---|
| Source | *df* | *MS* | *F* | Cohen's *f* | *MS* | *F* | Cohen's *f* |
| Group (G) | 4 | 12,845.73 | 3.82** | 0.47 | 432.73 | 0.26 | 0.12 |
| Error: G | 70 | 3,367.19 | | | 1,698.44 | | |
| Risk (R) | 2 | 20,578.87 | 53.46** | 0.87 | 12,285.23 | 51.00** | 0.85 |
| G × R | 8 | 652.99 | 1.70 | 0.31 | 654.80 | 2.72** | 0.39 |
| Error: R | 140 | 384.96 | | | 240.88 | | |
| Replication (r) | 2 | 2,463.27 | 11.88** | 0.41 | 3,207.68 | 14.36** | 0.45 |
| G × r | 8 | 290.28 | 1.40 | 0.28 | 176.57 | 0.79 | 0.21 |
| Error: r | 140 | 207.27 | | | 223.41 | | |
| R × r | 4 | 7,430.11 | 40.73** | 0.76 | 16,644.08 | 69.85** | 1.00 |
| G × R × r | 16 | 146.58 | 0.80 | 0.21 | 229.30 | 0.96 | 0.23 |
| Error: R × r | 280 | 182.41 | | | 238.30 | | |

*Note.* MI = myocardial infarction (heart attack); CAD = coronary artery disease.
** $p < .01$.

*Differences in risk tolerances across knowledge groups.* Regression analyses presented above indicated that risk tolerances contributed to decision making beyond the effects of risk perception and that tolerance for MI risk and for CAD probability each contributed uniquely. Thus, we performed an ANOVA on participants' risk tolerances for MI risk and CAD probability (as repeated measures), comparing the five knowledge groups. Patient risk and replications were not factors because all nine patients were used to derive two scores for risk tolerance, one for MI and one for CAD. Both of the main effects were significant, as was their interaction. Risk tolerance was lower for MI risk than for CAD probability (see Table 7). Knowledge group was also significant (see Table 7). Family practice physicians had lower tolerance for both kinds of risk than did more knowledgeable physicians; the maximum probability on average at which family practice physicians discharged patients was only 22% (*SE* = 4.93), whereas other physicians tolerated risks that were more than 20% higher, 41% (*SE* = 3.82), 45% (*SE* = 3.72), and 41% (*SE* = 4.73) for emergency medicine physicians, internal medicine physicians, and cardiologists, respectively. Students had a higher tolerance for risk (at 48%, *SE* = 5.69) than did family practice physicians. Thus, consistent with our hypothesis, physicians differing in knowledge differed in risk tolerance such that lower knowledge was associated with lower risk tolerance.

Table 7
*Analysis of Variance Results for Effects of Knowledge Group on Risk Tolerances for MI Risk and CAD Probability*

| Source | *df* | *MS* | *F* | Cohen's *f* |
|---|---|---|---|---|
| Group (G) | 4 | 2,561.12 | 4.40** | 0.50 |
| Error: G | 70 | 582.48 | | |
| Tolerance (T) | 1 | 33,143.60 | 110.07** | 1.25 |
| G × T | 4 | 1,286.83 | 4.27** | 0.49 |
| Error: T | 70 | 301.11 | | |

*Note.* MI = myocardial infarction (heart attack); CAD = coronary artery disease.
** $p < .01$.

Knowledge group also interacted significantly with risk tolerance (see Table 7). As can be seen in Figure 2, although tolerance of risk for MI was lower than that for CAD overall, the difference between these risk tolerances was not constant across knowledge groups: Differences were 19% for medical students, 26% for family practice physicians, 22% for emergency medicine physicians, 36% for internal medicine physicians, and 53% for cardiologists. Thus, cardiologists discriminated sharply (as did, to a lesser extent, internal medicine and other physicians) between the imminent risk of MI, for which they had a very low tolerance, and the probability of CAD, for which they had a very high tolerance (despite the admonitions of the guidelines that CAD probability is relevant to admission decisions). Indeed, risk tolerance for CAD was so high among cardiologists that the maximum probability at which they would discharge patients approached the judged probability of CAD even for the higher risk patients. Hence, MI risk and MI risk tolerance mainly determined triage decisions for the highest knowledge group, rather than CAD probability and risk tolerance for CAD. Overall, knowledge groups differed in discrimination between risk tolerances such that higher knowledge groups were less tolerant of risk on one dimension (MI) than they were on the other (CAD).

Risk tolerance was not significantly related to physicians' attitudes about different decision errors (i.e., their ratings of the relative undesirability of failure to admit and unnecessary admission). However, the range of attitudes was restricted: Over 95% of physicians indicated that failure to admit a patient who should have been admitted was worse than unnecessary admission, although this view was held to different degrees (on the evidence of 0%–100% ratings of those errors). Such ratings were not predictive of a bias to admit low-risk patients. For example, despite rating failure to admit as highly undesirable, cardiologists were least likely to admit low-risk patients. Family practice physicians, despite rating failure to admit as more acceptable than did any other group, were most likely to admit low-risk patients.

*Hedging by selecting the middle category.* The last analysis we report concerning risk tolerance and uncertainty concerns hedging—that is, selecting the middle response category. Selecting

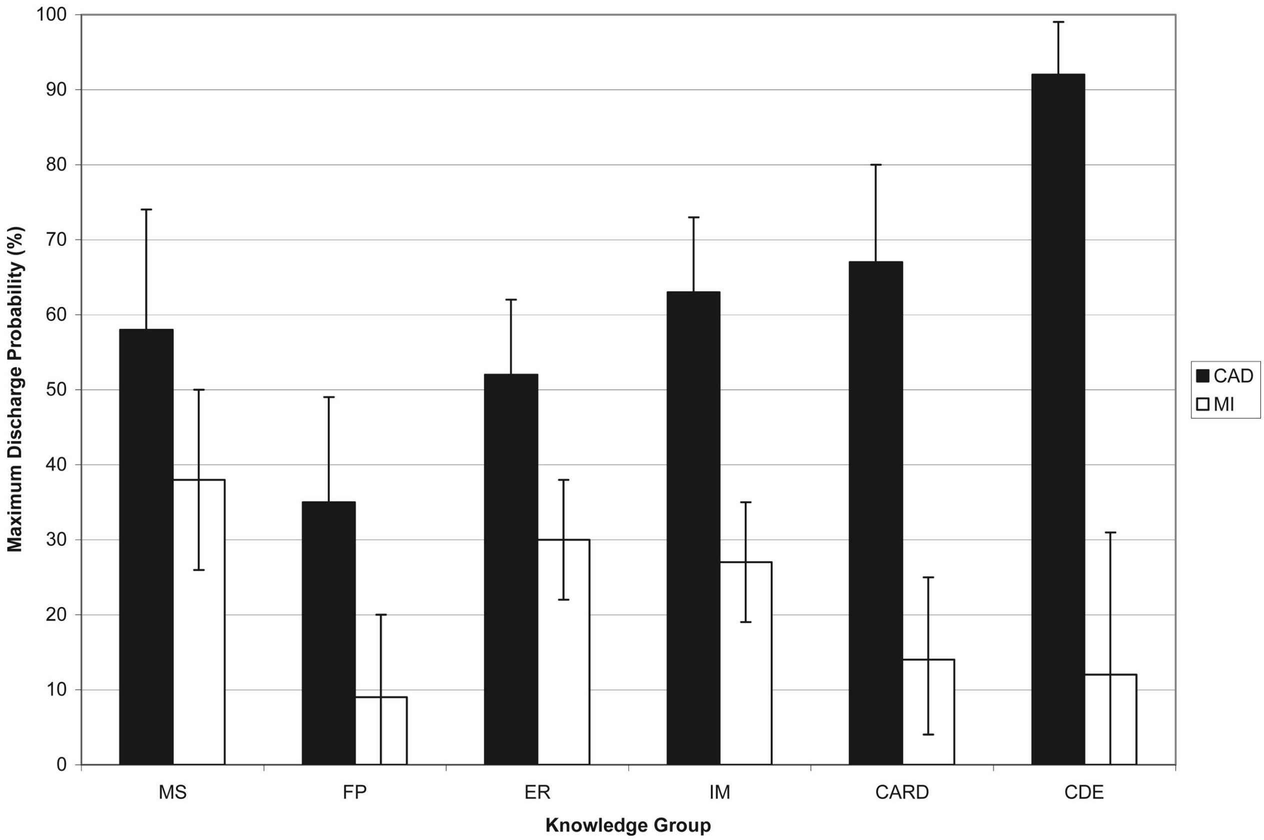


*Figure 2.* Mean tolerance for risk (maximum probability at which a discharge decision was made across all patients) for coronary artery disease (CAD) and myocardial infarction (MI) probability judgments, presented separately for six groups differing in level of knowledge (MS = medical students, FP = family practice physicians, ER = emergency medicine physicians, IM = internal medicine physicians, CARD = cardiologists, CDE = cardiology experts). Error bars represent 95% confidence intervals.

a middle response (e.g., admitting a patient but not to intensive care) minimizes the degree to which a decision could be wrong. At most, the middle category is one decision category away from the correct one, as opposed to discharging a patient who belongs in intensive care or admitting a patient to intensive care who should be discharged (both of which are two categories off from the correct one). In the absence of knowledge about the correct decision, selecting the middle category is the best bet to minimize risks. Admitting a patient to a ward bed avoids the risk of discharging a patient who could have an MI and die, and it also avoids the possibility of committing expensive resources on the basis of an uncertain judgment of risk. Thus, we compared the total number of times participants selected the middle decision category, regardless of patient risk, across knowledge groups. The latter factor was significant, $F(4, 70) = 4.64$, $MSE = 4.77$, $p < .01$, $f = 0.52$. Family practice physicians chose the middle category more often ($M = 7.67$, $SE = 0.63$) than did emergency medicine physicians ($M = 5.75$, $SE = 1.86$), internal medicine physicians ($M = 4.67$, $SE = 0.48$), or cardiologists ($M = 4.38$, $SE = 0.61$). Thus, our hypothesis that lower knowledge groups would choose the middle category more often was confirmed.

## *Internal Coherence and Cognitive Processing*

Internal coherence was assessed in three ways: (a) overall consistency between judgments and decisions, (b) correlations between judgments and decisions within each knowledge group, and (c) internal coherence as measured by disjunctive errors in probability judgments (i.e., assessment of whether participants' disjunctive probability judgments of MI or CAD risk were inconsistent with their MI or CAD judgments).

*Consistency between judgments and decisions.* First, as reported above, regression analyses for each category of patient risk indicated that judgments of MI risk and CAD probability were significantly related to triage decisions (see Table 4). This pattern reflects internal coherence, because patients judged to be at higher risk were triaged to higher levels of care (analogous to people being willing to pay higher rents for apartments that they rate as more desirable; see Fischer & Hawkins, 1993). Overall, as can be seen in Table 4, decisions were more consistently related to MI risk than they were to CAD probability.

*Correlations between judgments and decisions within knowledge groups.* Second, for each knowledge group, we averaged estimates for MI risk and for CAD probability for each patient, and we correlated these estimates with level of triage and with admission probability for each patient using Spearman's rho. The resulting correlations are reported in Table 8. Because these correlations were computed across patients rather than across participants, each knowledge group had identical degrees of freedom. That is, for each group, nine pairs of means were correlated for admission probability, and nine pairs of means were correlated for triage decision. The general pattern of significance was that both CAD probability and MI risk were strongly related to admission probability and to triage decision. However, for the highest knowledge groups, cardiologists and expert cardiologists, only the dimension of MI risk was significantly related to admission probability and to triage decision. The one exception to this pattern occurred for family practice physicians in the triage analysis and was likely attributable to restriction of range in their decisions; as demonstrated in the hedging analysis, family practice physicians selected Level 2 for almost all of their triage decisions, leaving little variability in their responses that could correlate with other measures. Nevertheless, even for family practice physicians, both dimensions of risk predicted admission probability, a pattern shared by all but the most expert decision makers. Correlations computed within each participant and then averaged showed the same pattern: Admission probability and admission decisions were related to both MI risk and CAD probability, except within the most knowledgeable groups, whose members relied on one dimension.

*Coherence as measured by disjunctive errors in probability judgments.* We assessed disjunctive errors by categorizing relations among judgments according to whether the relations violated basic ordinal properties of probabilities. For each participant, *MI or CAD* judgments were categorized for each patient in terms of their relations to the participant's separate CAD and MI estimates for that patient—that is, whether they satisfied the constraint that "or" probabilities should be equal to or greater than each of their component probabilities. If the disjunctive judgment was smaller than either component probability, an error was scored. We used ANOVAs to compare error rates for knowledge groups, level of risk, and patient replication. Overall, 22% of disjunctive judgments violated logical constraints. However, the analysis did not yield any significant effects: The distribution of violations did not vary by knowledge group, $F(4, 70) = 0.59$, $MSE = 0.68$, *ns*; level of patient risk, $F(2, 140) = 2.92$, $MSE = 0.13$, *ns*; replication, $F(2, 140) = 0.34$, $MSE = 0.13$, *ns*; or resident versus attending physician status (as reported above).

### Comparisons With Experts

Because large deviations from the guidelines were obtained (i.e., agreement was only 46%, and effect sizes were large), participants' judgments and decisions were also compared with those of 3 physicians who are nationally recognized experts in the diagnosis and triage of unstable angina. These experts responded to the same patient descriptions.

*Experts' admission decisions and probabilities.* Decisions were unanimous, except for two patients; for them, 2 of the 3 experts concurred. Table 9 shows agreement (strict agreement and admission agreement) between the experts and the guidelines. Decisions for each patient can be inferred by recalling that strict agreement for low-risk patients means the decision was discharge; similarly, for high-risk patients, strict agreement was admission to cardiac intensive care; and the lack of either strict or admission agreement for intermediate patients meant that they were discharged. As indicated in Table 9, experts mainly discharged patients for outpatient follow-up or placed them in intensive care. Note that all of the intermediate-risk patients were discharged by all three experts. The middle decision category was rarely used. Although the small sample size of national experts precluded their inclusion in ANOVAs as an additional knowledge group, their results are shown for illustration in Figures 1 and 2.

*Consistency, correlations, and coherence.* As can be seen in Table 9, top experts' triage decisions and admission probabilities were strongly related to risk of MI but not to CAD probability. The asymmetry in these relations was more extreme when compared with the larger sample. Computing correlations within participants and averaging them yielded the same pattern: for admission probability, Spearman's $\rho(7) = .99$, $p < .01$ (for MI) versus $\rho(7) = .15$, *ns* (for CAD); for triage decisions, $\rho(7) = .91$, $p < .05$ (for MI) versus $\rho(7) = .26$, *ns* (for CAD). Thus, national experts based their decisions almost exclusively on the single dimension of imminent risk of MI rather than the less immediately relevant probability of underlying CAD. As in the larger sample, however, internal co-

Table 8
*Correlations Between Estimates of MI Risk and CAD Probability With Admission Probability and Triage Decision for Groups Differing in Knowledge of Cardiovascular Disease*

| Knowledge group | Admission probability | | Triage decision | |
|---|---|---|---|---|
| | CAD | MI | CAD | MI |
| Medical students | .85** | .88** | .78** | .85** |
| Family practice physicians | .68* | .95** | .57 | .96** |
| Emergency medicine physicians | .92** | .97** | .73** | .82** |
| Internal medicine physicians | .87** | .93** | .63 | .95** |
| Cardiologists | .62 | .93** | .63 | .95** |
| Cardiology experts | .10 | .83** | .34 | .95** |

*Note.* Correlations are Spearman's rhos. For each group, nine pairs of means were correlated for both admission probability and triage decision. MI = myocardial infarction (heart attack); CAD = coronary artery disease.
* $p < .05$. ** $p < .01$.

Table 9
*Proportions of Strict Agreement and Admission Agreement Between Expert Cardiologists and the Guidelines for Patients at Low, Intermediate, and High Levels of Risk*

| Risk level and patient | Strict agreement | Admission agreement |
|---|---|---|
| Low | | |
| 1 | 1.00 | 1.00 |
| 2 | 1.00 | 1.00 |
| 3 | .00 | .00 |
| Intermediate | | |
| 4 | .00 | .00 |
| 5 | .00 | .00 |
| 6 | .00 | .00 |
| High | | |
| 7 | .67 | 1.00 |
| 8 | 1.00 | 1.00 |
| 9 | 1.00 | 1.00 |

*Note.* Strict agreement was based on the guidelines for unstable angina. Admission agreement was based on assessment of either admission decision (ward bed or cardiac intensive care unit) as agreement for intermediate- and high-risk patients.

herence was violated for a substantial percentage of "or" judgments: 30% of disjunctive judgments exhibited errors.

*Deviations from experts' recommendations.* We reanalyzed the larger sample's decisions using the experts' decisions in lieu of the guidelines. Goodness-of-fit tests indicated significant lack of fit between the decisions of the larger sample and those of the experts for all but one of the patients (Patient 3): for Patients 1–9, respectively, $t(74) = 7.05$, 9.70, 0.47 (*ns*), 16.52, 9.20, 14.22, 13.00, 8.51, and 10.54 (all $p\text{s} < .01$ except for Patient 3). The respective effects sizes, computed as Cohen's $d$, were 1.64, 2.25, 0.11, 3.84, 2.14, 3.31, 3.02, 1.98, and 2.45.

We also evaluated proportions of strict agreement and admission agreement, using the experts' decisions as a standard. As in the strict agreement and admission agreement analyses presented above, similar effects for patient risk, knowledge group, and interactions between patient risk and knowledge group were obtained (because of this similarity, these results are not reported in detail). Overall, 45% of physicians' decisions agreed with the experts' specific decisions, and 66% agreed with respect to admission versus nonadmission. Residents and attending physicians did not differ significantly in these analyses.

## Discussion

Professionals are called on to make important decisions about risk, sometimes with life-or-death consequences. Practice variation—when professionals treat similar cases or situations differently—is a recurring problem, one that is addressed through continuing education, publication of professional guidelines, and other quality assurance measures (e.g., Fuster, 1997). Medical error, pilot error, misidentification of nuclear threats (or other weapons of mass destruction), and other professional lapses in risky decision making are often preventable, in principle. For example, a 2000 Institute of Medicine report stated that the number of deaths attributable to *preventable* medical error in the United States is between 44,000 and 98,000 per year (Kohn et al., 2000; see also Hollnagel, 1993; Leape, 2000; Leape, Lawthers, Brennan, & Johnson, 1993). Preventing such errors requires a better understanding of their causes—namely, the psychological processes that underlie risky decision making.

This study addressed the following hypotheses: (a) that trained professionals would, despite their training, deviate from practice guidelines and from top experts' recommendations; (b) that more knowledgeable professionals would be better able to discriminate between low- and high-risk patients, as reflected in admission decisions and risk perceptions for MI and CAD; (c) that physicians differing in knowledge would also differ in risk tolerance such that lower knowledge was associated with lower risk tolerance and with hedging (choosing the middle category in admission decisions); (d) that these differences in risk tolerance would predict unique variance in decision making, beyond that accounted for by risk perceptions; (e) that professionals at higher levels of knowledge would process fewer dimensions of information to make decisions; (f) that the most advanced decision makers would make sharper all-or-none distinctions among decision categories; and (g) that even advanced decision makers would be subject to errors of coherence in disjunctive probability judgments, because such errors are a result of processing confusions rather than knowledge limitations. We now discuss each of these hypotheses in turn.

Regarding (a), we found that highly trained professionals deviated significantly from national guidelines for best practices. Physicians' decisions differed significantly from both the guidelines and from those of a panel of experts (suggesting that the observed practice variation was not a result of ambiguity associated with interpreting the guidelines). The guidelines had been disseminated to the professionals as part of their practice and had been discussed in special meetings. Commitment to the guidelines had been expressed by those attending the meetings. Nevertheless, the treatment and diagnosis of these, albeit hypothetical, patients did not meet national standards (see van Miltenburg-van Zijl, Bossuyt, Nette, Simoons, & Taylor, 1997). Similar disparities have been seen in treatment and diagnosis of actual patients, though because the real patients were not identical across physicians, it is always possible that regional, demographic, or other defensible differences explained practice variation (e.g., Wennberg, Freeman, & Culp, 1987). In this study, the patients were identical, making it impossible to explain differences in practice by appealing to differences in patients.

Regarding (b), the present data consistently showed that groups of students and physicians who differed in knowledge also differed in their medical decision making, illustrating practice variation even among highly trained professionals. The ability to discriminate between low- and high-risk patients also differed across physician groups for patient risk factors, such as the presence of CAD. Because of limited resources, most systems designed to identify risk have less well-trained people make initial assessments. Once some threshold of risk is identified, cases or situations are then passed on to those with greater expertise for further scrutiny. The gatekeeper system of managed care, in which generalist physicians make initial assessments and then refer cases to specialists if warranted, is such a system of screening and triage (e.g., Franks, Clancy, & Nutting, 1992; Sulmasy, 1993). However, although all of the physicians were "qualified" to make these decisions, our data indicate that specialists, those with greater domain-specific knowledge, are better able to discriminate low from high risk for subtle judgments such as unstable angina.

Regarding (c), practice variation is ordinarily assumed to be undesirable. However, our data illustrate that professionals predictably exhibit behavior that leads to desirable practice variation. That is, professionals with lower levels of knowledge (family practice physicians) were more likely to hedge in their decision making by choosing intermediate levels of care (protecting patients against more extreme errors in triage). Similarly, these professionals had lower tolerance of risk for both MI risk and CAD probability than did more knowledgeable physicians, which should protect against more dangerous, false negative errors (failing to admit a patient who should be admitted). If hedging is associated with lower levels of accuracy, as it was in this study, it is a desirable source of practice variation that should minimize human error.

Regarding (d), for all groups of participants, risk tolerances predicted decisions above and beyond risk perceptions (Swets et al., 2000). Unique variance was associated with tolerance of risk for MI as well as tolerance of probability of CAD. Risk tolerance in professional contexts has rarely been studied systematically, but prior work suggests that external factors (e.g., pressure to discharge patients to curb costs) as well as internal factors (e.g., individual differences in personality) are likely to influence tolerance for risk (e.g., Reyna & Farley, in press; Swets et al., 2000). Despite differences in degree of risk tolerance, participants overwhelmingly preferred false positive over false negative errors, and this asymmetry was also reflected in their greater agreement with the guidelines for high-risk patients (because high-risk patients were admitted). Erring on the side of caution, however, produces inevitable increases in false alarms, all other factors being equal, particularly without increases in knowledge.

To summarize our discussion so far, differences in knowledge were associated with differences in risk perception (and, hence, discrimination) and risk tolerance, which in turn predicted differences in decision making, accounting for practice variation. Despite the absence of scientifically precise information about exact risks and probabilities in the published literature, professionals could reliably discriminate low-risk from high-risk patients. As von Neumann and Morgenstern (1944) observed, humans are able to process vagueness, noise, and ambiguity in the signal stream more easily than computers. Regarding (e), according to fuzzy-trace theory, they are able to do so because human judgment and decision making is not fundamentally computational. Instead, humans have a fuzzy-processing preference that generally leads to superior decisions based on less information processed more simply (see also Dijksterhuis, Bos, Nordgren, & van Baaren, 2006), and this preference grows as decision makers develop and become more knowledgeable about a task. Hence, applying fuzzy-trace theory, we made the counterintuitive prediction that professionals at higher levels of knowledge would process fewer dimensions of information (Hypothesis e) and that the most advanced decision makers would make sharper all-or-none distinctions among decision categories (Hypothesis f). These theoretically motivated, but surprising, predictions were upheld. As can be seen in Table 8, both dimensions of risk predicted admission probability and admission decision at lower levels of knowledge, but the two most knowledgeable groups (cardiologists and nationally recognized experts in cardiology) only relied on one dimension of risk. Thus, experts achieved *better* discrimination by processing *less* information.

Figure 2 echoes the theme of better discrimination through the processing of less information (Hypothesis e). As level of knowledge increased, risk tolerance for CAD generally increased, whereas risk tolerance for MI decreased. The most knowledgeable professionals had the largest separation between risk tolerances, with the tolerance for CAD so high that CAD did not figure in most admission decisions. In other words, for cardiologists and expert cardiologists, tolerance for CAD was so high that the estimated probabilities of CAD even for high-risk patients usually did not exceed their threshold for admission. These experts were exquisitely sensitive to the risk of MI, however. Thus, knowledge groups differed in their discrimination between risk tolerances as well as risk perceptions, and more knowledgeable professionals relied mainly on one dimension of risk tolerance to make admission decisions.

Regarding (f), as Table 8 shows, the most expert physicians (top cardiologists) sharply distinguished patients who were at imminent risk of MI, and needed to be admitted to intensive care, from those who could be discharged with follow-up (even discharging patients at supposedly intermediate levels of risk). Expert cardiologists described patients as progressing along a pathophysiological course of CAD that takes a turn, placing them at risk of an MI, or continues to slowly progress, not placing them at imminent risk. This disposition of patients is sharply all-or-none—discharge or intensive care—as is the conceptualization of risk (see Reyna, Adam, Poirier, LeCroy, & Brainerd, 2005).

Regarding (g), despite consistent evidence that more knowledgeable professionals made superior decisions with respect to external criteria (e.g., they were better at discriminating risk according to the guidelines), they were not significantly superior with respect to measures of internal coherence, as predicted by fuzzy-trace theory. For example, disjunctive probability judgment errors did not differ statistically across knowledge groups. Errors ranged from about one out of five judgments to one out of three for top experts. Such disjunctive judgments are clinically relevant because they are supposed to be the basis for admission decisions. Physicians, like other professionals, are expected to integrate information about risks and probabilities to formulate overall judgments (e.g., regarding admission). For these probability integration tasks, such factors as limited memory capacity, conceptual deficits, and illogical reasoning have been ruled out as sources of performance error in student samples (Brainerd & Reyna, 1990; Reyna, 1992; Wolfe, 1995). However, researchers have argued that experts should be less vulnerable to errors in probability judgment than are students and other nonexperts (e.g., Dougherty, Gronlund, & Gettys, 2003). The current results do not support this conclusion for disjunctive probability judgments. Instead, the results are compatible with the conclusion that experts are vulnerable to interference effects created by overlapping or nested classes, and these effects are exacerbated by compelling gists of the classes that distract competent reasoners from correcting their processing errors (Brainerd & Reyna, 1990; Lloyd & Reyna, 2001b; Reyna, 1991; see also Sloman et al., 2003, for replication in a somewhat different task). Thus, the current results add to other recent findings illustrating dissociation in performance between coherence and correspondence criteria for rationality, supporting the emerging view that these criteria cannot be reduced to one another and that both criteria are essential for judging rationality (Doherty, 2003; Reyna & Farley, in press).

This study has a number of limitations, which should be explored in further research. First, although the use of hypothetical patients allows experimental control over patient characteristics and ensures that physicians are judged on a common metric, it would be informative to examine analogous decision making in real patients (comparisons between hypothetical and real patients have generally yielded similar results for the condition studied here [unstable angina]; e.g., van Miltenburg-van Zijl et al., 1997). Second, this study did not include a nationally representative sample of physicians in specialty groups; thus, broad conclusions about such groups are not warranted on the basis of our data. In fact, a key comparison would be to examine risk discrimination and gist-based information processing for physicians varying in domain-specific knowledge (explicitly assessed) within specialty groups and, ultimately, to study the impact of specific knowledge and experience using experimental designs that vary exposure to theoretically motivated educational interventions.

## References


Adam, M. B., & Reyna, V. F. (2005). Coherence and correspondence criteria for rationality: Experts' estimation of risks of sexually transmitted infections. *Journal of Behavioral Decision Making, 18,* 169–186.

Bauer, M. S., Callahan, A. M., Jampala, C., Petty, F., Sajatovic, M., Schaefer, V., et al. (1999). Clinical practice guidelines for bipolar disorder from the Department of Veterans Affairs. *Journal of Clinical Psychiatry, 60,* 9–21.

Brainerd, C. J., & Reyna, V. F. (1990). Inclusion illusions: Fuzzy-trace theory and perceptual salience effects in cognitive development. *Developmental Review, 10,* 365–403.

Braunwald, E., Antman, E. M., Beasley, J. W., Califf, R. M., Cheitlin, M. D., Hochman, J. S., et al. (2000). ACC/AHA guidelines for the management of patients with unstable angina and non-ST-segment elevation myocardial infarction: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on the Management of Patients With Unstable Angina). *Journal of the American College of Cardiology, 36,* 970–1062.

Braunwald, E., Antman, E. M., Beasley, J. W., Califf, R. M., Cheitlin, M. D., Hochman, J. S., et al. (2002). ACC/AHA 2002 guideline update for the management of patients with unstable angina and non-ST-segment elevation myocardial infarction—Summary article: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology, 40,* 1366–1374.

Braunwald, E., Mark, D. B., Jones, R. H., Cheitlin, M. D., Fuster, V., & McCauley, K. M. (1994). *Unstable angina: Diagnosis and management* (Clinical Practice Guideline No. 10 [amended]; AHCPR Publication No. 94–0602). Rockville, MD: U.S. Department of Health and Human Services.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74,* 271–280.

Chin, M. H., Friedman, P. D., Cassel, C. K., & Lang, R. M. (1997). Differences in generalist and specialist physicians' knowledge and use of angiotensin-converting enzyme inhibitors for congestive heart failure. *Journal of General Internal Medicine, 12,* 523–530.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Collinson, P. O., Premachandram, S., & Hashemi, K. (2000). Prospective audit of incidence of prognostically important myocardial damage in patients discharged from emergency department. *British Medical Journal, 320,* 1702–1705.

Dawes, R. M. (1988). *Rational choice in an uncertain world.* San Diego, CA: Harcourt Brace Jovanovich.

Depression Guideline Panel. (1993). Depression in primary care, Vol. 2: Treatment of major depression (Clinical Practice Guideline, No. 5; DHHS Publication No. AHCPR 93–0551). Washington, DC: U.S. Government Printing Office.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006, February 17). On making the right choice: The deliberation-without-attention effect. *Science, 311,* 1005–1007.

Doherty, M. E. (2003). Optimists, pessimists, and realists. In S. L. Schnieder & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 643–679). New York: Cambridge University Press.

Dougherty, M. R. P., Gronlund, S. D., & Gettys, C. F. (2003). Memory as a fundamental heuristic for decision making. In S. L. Schnieder & J. Shanteau (Eds.), *Emerging perspectives on judgments and decision research* (pp. 125–164). New York: Cambridge University Press.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). New York: Cambridge University Press.

Eddy, D. M. (1996). *Clinical decision making: From theory to practice.* Sudbury, MA: Jones & Bartlett.

Federal Aviation Administration. (1994). *Takeoff safety training aid* (Reference No. PB93780013). Springfield, VA: National Technical Information Service.

Fischer, G. W., & Hawkins, S. A. (1993). Strategy compatibility, scale compatibility, and the prominence effect. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 580–597.

Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S., & Keeney, R. (1981). *Acceptable risk.* New York: Cambridge University Press.

Fitts, P. M. (1947). Psychological research on equipment designs in the AAF. *American Psychologist, 2,* 93–98.

Franks, P., Clancy, C. M., & Nutting, P. A. (1992, August 6). Gatekeeping revisited—Protecting patients from overtreatment. *New England Journal of Medicine, 327,* 424–429.

Fuster, V. (1997). The generalist/cardiovascular specialist: A proposal for a new training track. *Annals of Internal Medicine, 127,* 630–634.

Graves, E. J., & Gillum, B. S. (1996). *1994 summary: National hospital discharge survey* (Advance Data No. 278). Hyattsville, MD: National Center for Health Statistics.

Harrold, L. R., Field, T. S., & Gurwitz, J. H. (1999). Knowledge, patterns of care, and outcomes of care for generalists and specialists. *Journal of General Internal Medicine, 14,* 499–511.

Hollnagel. E. (1993). *Human reliability analysis: Context and control.* London: Academic Press.

Johnson, B. B. (1993). Advancing understanding of knowledge's role in lay risk perception. *RISK: Issues in Health and Safety, 4,* 189–212.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press

Katz, D., Griffith, J., Beshansky, J., & Selker, H. (1996). The use of empiric clinical data in the evaluation of practice guidelines for unstable angina. *Journal of the American Medical Association, 276,* 1568–1574.

Klaczynski, P. (2005). Metacognition and cognitive variability: A dual-process model of decision making and its development. In J. Jacobs & P. Klaczynski (Eds.), *The development of judgment and decision making in children and adolescents* (pp. 39–76). Mahwah, NJ: Erlbaum.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (2000). *To err is*

*human: Building a safer health system.* Washington, DC: National Academies Press.
Leape, L. L. (2000). Institute of Medicine medical error figures are not exaggerated. *Journal of the American Medical Association, 284,* 95–97.
Leape, L. L., Lawthers, A. G., Brennan, T. A., & Johnson, W. G. (1993). Preventing medical injury. *Quality Review Bulletin, 19,* 144–149.
Lee, T. H., Pearson, S. D., & Johnson, P. A. (1995). Failure of information as an intervention to modify clinical management. *Annals of Internal Medicine, 122,* 434–437.
Lloyd, F. J., & Reyna, V. F. (2001a). Clinical decision making and information management in the era of managed care. In R. C. Becker & J. S. Alpert (Eds.), *Cardiovascular medicine: Practice and management* (pp. 719–732). New York: Arnold.
Lloyd, F. J., & Reyna, V. F. (2001b). A Web exercise in evidence-based medicine using cognitive theory. *Journal of General Internal Medicine, 16,* 94–99.
Loewenstein, G., & Mather, J. (1990). Dynamic processes in risk perception. *Journal of Risk and Uncertainty, 3,* 155–175.
Nightingale, S. D. (1987). Risk preference and laboratory test selection. *Journal of General Internal Medicine, 2,* 25–28.
Poses, R. M., Wigton, R. S., Cebul, R. D., Centor, R. M., Collins, M., & Fleischli, G. J. (1993). Practice variation for patients with pharyngitis: The importance of variability in patients' clinical characteristics and in physicians' response to them. *Medical Decision Making, 13,* 293–301.
Reason, J. (1990). *Human error.* New York: Cambridge University Press.
Reyna, V. F. (1991). Class inclusion, the conjunction fallacy, and other cognitive illusions. *Developmental Review, 11,* 317–336.
Reyna, V. F. (1992). Reasoning, remembering, and their relationship: Social, cognitive, and developmental issues. In M. L. Howe, C. J. Brainerd, & V. F. Reyna (Eds.), *Development of long-term retention* (pp. 103–127). New York: Springer-Verlag.
Reyna, V. F. (1995). Interference effects in memory and reasoning: A fuzzy-trace theory analysis. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 29–59). San Diego, CA: Academic Press.
Reyna, V. F. (2004). How people make decisions that involve risk: A dual-processes approach. *Current Directions in Psychological Science, 13,* 60–66.
Reyna, V. F., & Adam, M. B. (2003). Fuzzy-trace theory, risk communication, and product labeling in sexually transmitted diseases. *Risk Analysis, 23,* 325–342.
Reyna, V. F., Adam, M. B., Poirier, K., LeCroy, C. W., & Brainerd, C. J. (2005). Risky decision-making in childhood and adolescence: A fuzzy-trace theory approach. In J. Jacobs & P. Klacynski (Eds.), *The development of judgment and decision making in children and adolescents* (pp. 77–106). Mahwah, NJ: Erlbaum.
Reyna, V. F., & Brainerd, C. J. (1991). Fuzzy-trace theory and the acquisition of scientific and mathematical concepts. *Learning and Individual Differences, 3,* 27–60.
Reyna, V. F., & Brainerd, C. J. (1993). Fuzzy memory and mathematics in the classroom. In G. M. Davies & R. H. Logie (Eds.), *Memory in everyday life* (pp. 91–119). Amsterdam: North-Holland.
Reyna, V. F., & Brainerd, C. J. (1994). The origins of probability judgment: A review of data and theories. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 239–272). New York: Wiley.
Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7,* 1–75.
Reyna, V. F., & Ellis, S. C. (1994). Fuzzy-trace theory and framing effects in children's risky decision making. *Psychological Science, 5,* 275–279.
Reyna, V. F., & Farley, F. (in press). Risk and rationality in adolescent decision making: Implications for theory, practice, and public policy. *Psychological Science in the Public Interest.*
Reyna, V. F., & Lloyd, F. J., & Brainerd, C. J. (2003). Memory, development, and rationality: An integrative theory of judgment and decision making. In S. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 201–245). New York: Cambridge University Press.
Salerno, S. M., Alguire, P. C., & Waxman, H. S. (2003). Competency in interpretation of 12-lead electrocardiograms: A summary and appraisal of published evidence. *Annals of Internal Medicine, 138,* 751–760.
Seeger, P., & Gabrielsson, A. (1968). Applicability of the Cochran *Q* test and the *F* test for statistical analysis of dichotomous data for dependent samples. *Psychological Bulletin, 69,* 269–277.
Shanteau, J. (1969). Psychological characteristics and strategies of expert decision makers. *Acta Psychologica, 68,* 203–215.
Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes, 91,* 296–309.
Slovic, P., Malmfors, T., Krewski, D., & Mertz, C. K. (1995). Intuitive toxicology: II. Expert and lay judgments of chemical risks in Canada. *Risk Analysis, 15,* 661–675.
Smith, J. F., & Kida, T. (1991). Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin, 109,* 472–485.
Streuber, S. D., Amsterdam, E. A., & Stebbins, C. L. (2006). Heart rate recovery in heart failure patients after a 12-week cardiac rehabilitation program. *American Journal of Cardiology, 97,* 694–698.
Sulmasy, D. (1993). Who keeps the gate? *Journal of the American Medical Association, 270,* 2116–2117.
Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist, 47,* 522–532.
Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1,* 1–26.
Taylor, T. R. (2000). Understanding the choices that patients make. *Journal of the American Board of Family Practice, 13,* 124–133.
Tierney, W. M., Overhage, J. M., Takesue, B., Harris, L., & Murray, M. (1995). Computerizing guidelines to improve care and patient outcomes: The example of heart failure. *Journal of the American Medical Informatics Association, 2,* 316–322.
Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101,* 547–567.
U.S. General Accounting Office. (2001, August). *General aviation: Status of the industry, related infrastructure, and safety issues* (Publication No. GAO-01–916). Washington, DC: Author.
van Miltenburg-van Zijl, A. J. M., Bossuyt, P. M. M., Nette, R. W., Simoons, M. L., & Taylor, T. R. (1997). Cardiologists' use of clinical information for management decisions for patients with unstable angina: A policy analysis. *Medical Decision Making, 17,* 292–297.
von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.
von Winterfeldt, & Edwards, W. (1986). *Decision analysis and behavioral research.* Cambridge, England: Cambridge University Press.
Weingarten, S. R. (1997). Practice guidelines and prediction rules should be subject to careful clinical testing. *Journal of the American Medical Association, 277,* 1977–1978.
Wennberg, J. E. (1987). Population illness rates do not explain population hospitalization rates. *Medical Care, 25,* 354–359.
Wennberg, J. E., Freeman, J. L., & Culp, W. J. (1987). Are hospital services rationed in New Haven or over-utilised in Boston? *The Lancet,* 1185–1188.
Wolfe, C. R. (1995). Information seeking on Bayesian conditional probability problems: A fuzzy-trace theory account. *Journal of Behavioral Decision Making, 8,* 85–108.
Yates, J. F. (1990). *Judgment and decision making.* Englewood Cliffs, NJ: Prentice Hall.

# Appendix

## One of Nine Case Descriptions

### History

A 65-year-old female comes to the emergency room; she has experienced increasing chest pressure and diaphoresis for 45 minutes. Over the past 2 days she experienced 2 mild episodes of this discomfort and has felt more tired than usual. There is no history of heart disease, but she was a heavy smoker until 1 year ago and has had long-standing hypertension currently controlled on medications. She denies significant previous illness except for a duodenal ulcer 3 years ago.

### Examination

An anxious, pale middle-aged woman with blood pressure of 154/90 and heart rate of 84. There is no jugular venous distension, and the carotid pulsations are normal. The lungs reveal basilar crackles, and there is an S4 gallop. The remainder of the examination is normal.

### Additional Data

The lab work, chest x-ray and ECG are normal.

### Questions

For the questions below, estimate risk and probability by using a 0–100% scale. Note that 0 means "no chance at all," 50 means "as likely as not" and 100 means "absolutely certain."

1. What is your estimate of the probability that this patient has coronary artery disease (CAD)? ________
2. What is your estimate of this patient's imminent risk of myocardial infarction (MI)? ________
3. What is your estimate of the probability that this patient has a condition *other than* CAD or risk of MI that is serious enough to warrant admission? ________
4. The decision to admit a patient does not necessarily indicate *absolute certainty* about the patient's condition. Regardless of your decision in this case, what is your estimate of the probability that this patient has (or will imminently develop) a condition serious enough to warrant admission? ________
5. What is your estimate of the probability that this patient has CAD or is at imminent risk of MI? ________
6. What is your triage decision in this case?

   a. Out-patient follow-up

   b. Admit to unmonitored bed (with Medicine Team)

   c. Admit to telemetry unit

   d. Admit to Cardiovascular Intensive Care Unit

   e. Other ________________ (please indicate where)

Received May 17, 2005
Revision received June 27, 2006
Accepted June 28, 2006 ■

**E-Mail Notification of Your Latest Issue Online!**

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at http://watson.apa.org/notify/ and you will be notified by e-mail when issues of interest to you become available!