Bayes theorem: Fully informed rational estimates of diagnostic probabilities

The following resources related to this article are available online at jada.ada.org (this information is current as of June 1, 2010):

Updated information and services including high-resolution figures, can be found in the online version of this article at:
http://jada.ada.org/cgi/content/full/141/6/658

Information about obtaining reprints of this article or about permission to reproduce this article in whole or in part can be found at:
http://www.ada.org/prof/resources/pubs/jada/permissions.asp
The probability that a diagnostic observation or test result is positive when a patient has the condition in question is not necessarily the same as the probability that the same patient has the condition if the observation or test result is positive. In the first case, the condition causes the evidence to appear; in the other case, the evidence is used to infer the existence of the underlying condition (the diagnosis). The former is a principal concern of the research community and those who develop tests; the latter is the primary concern of practitioners.

Researchers have studied extensively the way in which physicians form judgments in clinical diagnosis.1-4 This study represents one of the first attempts to understand how that process takes place in dentistry. The most straightforward view of diagnosis is that practitioners directly observe the condition that requires treatment, with no intervening process of inference connecting evidence to the background condition; practitioners simply understand what needs to be done. A more nuanced interpretation is that practitioners observe evidence directly and use the evidence as a basis for making correct diagnoses; however, the process they use in proceeding from evi-

**ABSTRACT**

**Background.** Research in medicine has shown that physicians have difficulty estimating the probability that a patient has a condition on the basis of available diagnostic evidence. They consistently undervalue baseline information about the patient relative to test information and are poor intuitive calculators of probability. The authors could not locate in the literature any studies of diagnostic probability estimates from baseline information and test data for dentists.

**Methods.** Using two vignettes that contained different baseline information, dental students and clinical faculty members estimated the probability that the described hypothetical patient had the condition in question. Respondents also commented on the project.

**Conclusions.** Both groups of respondents overemphasized the importance of test evidence relative to baseline information, although experienced practitioners did so to a lesser extent than did students. Respondents, especially practitioners, expressed resistance to performing a diagnostic task that required precise estimates of probability.

**Clinical Implications.** Dentists appear to estimate diagnostic probabilities in an intuitive fashion, but they do so imprecisely. Clinical experience provides some protection against the bias of overestimating test evidence compared with baseline information. These findings raise questions about how practitioners use probability estimates and whether other models also may play a role. The incorporation of information from evidence-based dentistry into practice requires better understanding.

**Key Words.** Clinical diagnosis; Bayes theorem; evidence-based dentistry; test sensitivity.

Dr. Chambers is a professor of dental education, Department of Dental Practice, Arthur A. Dugoni School of Dentistry, University of the Pacific, San Francisco. Contact Dr. Chambers at 19800 7th St. E., Sonoma, Calif. 95476, e-mail “dchambers@pacific.edu”.

Mr. Mirchel is a third-year dental student, Arthur A. Dugoni School of Dentistry, University of the Pacific, San Francisco.

Dr. Lundergan is a professor of periodontics, Arthur A. Dugoni School of Dentistry, University of the Pacific, San Francisco.
The decision maker uses the baseline likelihood of the condition to weight the probability that a finding is a clear signal that the underlying condition is present. The second factor that must be considered is the quality of the observation. Intuitively, a poorly exposed radiograph more likely represents false-positive evidence than does a well-exposed radiograph. Researchers and clinicians must adjust the evidence, whether positive or negative, for its sensitivity (that is, the chance of a positive observation when the condition exists) and for the baseline likelihood of the condition.

The theoretically correct method for estimating diagnostic probabilities involves the use of Bayes theorem. The sidebar to this article provides an example that describes this approach. The concept of Bayes theorem is to consider all the ways in which a test or observation can produce a positive finding. An obvious example is when a true-positive result occurs because the test or observation is sensitive. Adjusting for the baseline likelihood of the condition provides the likelihood that acting on the positive test result or observation will be appropriate.

However, there is a second way in which findings can appear to indicate the presence of a condition: false-positive test results or observations can occur, and these must be adjusted for the likelihood that the condition does not exist. (By definition, false-positive findings exist only in patients who do not have the condition.) According to Bayes theorem, the chance of being correct in acting on a positive test result or observation is the ratio of the correct positive finding to the total (both correct and incorrect) positive findings.

Eddy and Lyman and Balducci conducted research on diagnostic decision making and reported large variability among physicians given a common set of information. According to other researchers, physicians’ estimates often are inaccurate when compared with fully informed estimates. Three general areas of difficulty in performing this decision-making task are described in the literature:

- difficulty in understanding the evidence provided by tests;
- underuse of baseline information about the patient;
- inappropriate integration of the available data.

The American College of Dentists provides a detailed review of the medical literature regarding the assessment of diagnostic probabilities and an explanation of the use of Bayes theorem in combining these data.

On the basis of our assessment of the literature regarding the accuracy of diagnostic probability estimates given baseline data and test evidence, we generated the following hypotheses for testing among dental students and clinical faculty members by using vignette simulations of a
Bayes theorem

Fully informed rational estimates of diagnostic probabilities

Recent discussions of changes in guidelines for prostate-specific antigen tests and mammography have focused attention on the fact that positive results from tests or from direct observation are not sufficient to establish the presence of a condition with 100 percent certainty. Two factors may temper positive diagnostic information. The first is the sensitivity of the test or observation; sometimes a false-positive result occurs. The second factor is the baseline prevalence of the condition.

A judgment that a common condition exists in a particular patient often is correct simply because the condition occurs so frequently in patients. The opposite is equally true; rare conditions remain somewhat unlikely, even when positive results from tests or direct observation occur.

The problem for clinicians is incorporating test sensitivity (freedom from false-positive observations) and baseline prevalence into a meaningful estimate of the likelihood that a patient has the condition in question. Health care professionals do this intuitively, with some degree of accuracy and possibly with some degree of bias. A precise way of accomplishing this is with Bayes theorem.1

Combining information about the sensitivity of an observation or a test result with the baseline prevalence of the condition in the population is somewhat complex. The rational approach is expressed in Bayes' theorem, which states that the likelihood that evidence in favor of a condition's existing is correct is the ratio of true-positive observations in the population divided by the probability of true-positive observations plus false-positive observations. In other words, how trustworthy is the observation or test result?

What we want to determine is the probability (Pr) that a condition (C) exists given the evidence (E): Pr(C|E). What we are given is the probability that the evidence exists given the condition Pr(E|C) and the baseline prevalence of the condition Pr(C). The fundamental concept in Bayes theorem is that there are two ways to obtain a positive test result or an involved a positive test result for a periodontal condition in a patient from a population group with a given prevalence of the condition; the test involved the use of an experimental piece of equipment. We provided specific quantitative values for the sensitivity of the test, the false-positive test results and the baseline prevalence of the condition in the population. We provided participants with all of the information needed to calculate precisely the likelihood of the patient's having the condition, and we did not provide any extraneous information in the vignette.

We presented two versions of the vignette to the groups; they differed only in the information contained regarding the baseline prevalence of the condition in the population. One version contained baseline information designed to raise the correct probability estimate above the test sensitivity value. In this version, sensitivity equals
observation; a positive result can occur when the condition actually exists (that is, a true-positive observation weighted by the baseline prevalence), and a positive result can occur when the condition does not exist (that is, a false-positive observation weighted by the baseline prevalence). The probability of being fooled by a false-positive observation can be calculated from the information in hand:

\[ \Pr(E \mid \neg C) = 1 - \Pr(E \mid C) \text{ and } \Pr(\neg C) = 1 - \Pr(C) \]

and Bayes theorem states

\[ \Pr(C \mid E) = \frac{\Pr(C) \cdot \Pr(E \mid C)}{\Pr(C) \cdot \Pr(E \mid C) + \Pr(\neg C) \cdot \Pr(E \mid \neg C)} \]

where \( \Pr \) equals probability, \( E \) equals the evidence, \( C \) equals the patient's having the condition, \( \neg C \) equals the patient's not having the condition.

Consider this example based on the vignettes used in this study. If a test with a sensitivity of 0.70 (that is, seven of 10 times the observation is correct) is applied to a patient with only a 10 percent chance of having the condition, the likelihood of observing a positive test result in this or any patient will be 0.07 (0.70 \times 0.10). The likelihood of observing a positive result in a patient who does not have the condition actually will be larger in this example. The likelihood of being misled by the test result is 0.27 (that is, 0.30—the chance of a false-positive result—or 1.0 – 0.70 multiplied by 0.90—the chance that the patient does not have the condition—or 1.00 – 0.10). The chance of observing a true- or false-positive test result or clinical observation is 0.07 (that is, a positive result for patients with the condition) plus 0.27 (that is, a positive result for patients without the condition) equals 0.34. The probability that a positive observation will accurately reveal the presence of the condition in question is 0.07 (true-positive test results) divided by 0.34 (all positive test results) equals 0.21.

Although the Bayesian calculation appears complex when explained in detail, it gives proper weight to the relevant factors in practitioners' use of clinical judgment and test results to estimate diagnostic probabilities. It is no more complex than the calculations drivers engage in when calculating the timing and force of braking when driving.

---


---

**ABBREVIATION KEY.** \( \Pr(C) \): Probability that a condition exists. \( \Pr(\neg C) \): Probability that a condition does not exist. \( \Pr(C \mid E) \): Probability that a condition exists in a patient given evidence of its existence. \( \Pr(E \mid C) \): Probability of a positive test result or observation in a patient who has a condition. \( \Pr(E \mid \neg C) \): Probability of a positive test result or observation in a patient who does not have a condition.

---

0.70, the false-positive rate equals 0.30 and the baseline prevalence of the condition equals 0.80. Given these data, the Bayesian diagnostic probability is 0.90; this vignette is referred to as the “90 percent vignette.” The second vignette was identical, except that the baseline prevalence of the condition equals 0.10. This difference has the effect of shifting the correct diagnosis in the opposite direction of the positive test results. These data combine to produce a Bayesian diagnostic probability of 0.21; thus, the second vignette is referred to as the “21 percent vignette.”

We presented the vignettes in 2008 to eight groups of first-year dental students as part of a course on clinical dental research (total class size = 143) and to five groups of clinical faculty members as part of a quarterly in-service training program (total attendance of approximately 80 people). The study protocol fell under the exempt classification for human research subjects, and participants were anonymous. Participation was voluntary. After the participants provided their estimates, we presented a 20-minute explanation of the Bayesian approach to diagnosis, which was illustrated by the stimulus vignettes.

We shuffled the stimulus vignettes before distributing them, and each respondent had an equal chance of receiving the 90 percent or the 21 percent vignette. We asked the students and faculty members to provide a single, quantitative estimate of the likelihood that the patient had the indicated condition on the basis of the available data, as well as to briefly describe the reasoning used. Many participants also provided comments. After making their individual estimates, the students worked in groups to develop a better understanding of the role that evidence plays in diagnosis; however, we did not gather any data from these discussions and
they did not affect the data reported in this study.

Using the participants’ descriptions of numerical estimates, we classified their responses according to a particular strategy or combination of strategies, as described in the Results section below. We set up the baseline values in the vignettes to be nonsymmetrical to ensure that each participant could arrive at his or her estimate only via a unique strategy. Some respondents volunteered comments on their response forms, and several expressed frustration that the vignettes contained insufficient information for them to make the requested estimate or that making any precise estimate was difficult or uncomfortable. We calculated a single proportion for the number of such comments divided by the number of responding students or clinical faculty members. One of us (D.W.C.) classified all responses according to strategy, and the other authors reviewed a sample of responses to confirm these classifications.

We tested the hypotheses by using the conventional t test for differences between proportions.

RESULTS

Thirteen students declined to participate or were absent on the day of the activity, and three handed in estimates that were indecipherable or implausible (for example, \( P > 1.0 \)). Approximately six faculty members declined to participate; five others handed in forms that contained only comments or unusable estimates. The effective sample size was 127 students (60 for the 90 percent vignette and 67 for the 21 percent vignette) and 69 faculty members (36 for the 90 percent vignette and 33 for the 21 percent vignette).

Figure 1 shows the distribution of the estimated likelihood of the patient’s having the periodontal condition for the 90 percent vignette, based on the evidence provided and baseline information. Figure 2 shows the distribution for the 21 percent vignette. For both vignettes, the estimates ranged from 0 to 100 percent. The standard deviations of the estimates are large relative to the means. Interquartile ranges for the 90 percent vignette were 0.60 to 0.70 for students and 0.60 to 0.80 for faculty members. For the 21 percent vignette, for which we expected the baseline data to lower the estimate, the interquartile range was 0.30 to 0.70 for students and 0.10 to 0.50 for faculty members. The distributions are not strictly continuous because each respondent’s chosen strategy yielded an exact probability. Nevertheless, the distributions are approximately normal.

Among faculty members, the hypothesized upward shift for the 90 percent vignette and the hypothesized downward shift for the 21 percent vignette appear in Figures 1 and 2, respectively, and in Table 1 (page 662). The greater accuracy of the estimated diagnostic probability among faculty members compared with that among students was significant at \( P = .048 \) for the 90 percent vignette and \( P = .008 \) for the 21 percent vignette. We reran the \( t \) test comparing the results for faculty members with those for students for the 21 percent vignette to exclude all estimates lower than 10 percent; this enabled us to eliminate the effects of the pronounced left-hand tail in the distribution (Figure 2). The results under this severely handicapped test still revealed a significantly greater use of baseline information by faculty members (\( P = .046 \)).

Diagnostic strategies. By considering both respondents’ numerical estimates and the reasons
provided in their comments, we were able to identify a diagnostic strategy for each respondent. We identified four approaches on the basis of a single source of information:

- the participant equated the likelihood of the patient’s having the condition with the baseline prevalence (while ignoring all other information);
- the participant set the diagnostic probability equal to the test sensitivity (while ignoring all other information);
- the participant equated the positive test result with a 100 percent likelihood of the patient’s having the condition (while ignoring all other information);
- the participant used the false-positive rate to determine a diagnosis (while ignoring all other information).

In addition to the above, we noted two strategies that made use of false-positive adjustments to the sensitivity rate. In one case, the respondent subtracted the absolute rate of false-positive observations from the test sensitivity rate (that is, 0.70 − 0.30 = 0.40). In the second case, the respondent subtracted the false-positive rate from the test sensitivity rate on a proportional basis (that is, 0.70 − (0.70 × 0.30) = 0.49). A related strategy involved subtracting the false-positive rate from the positive test result (that is, 1.00 − 0.30 = 0.70).

Finally, we identified three mixed strategies: one involved the use of various combinations of baseline and test sensitivity data; one involved subtracting baseline information from the test sensitivity rate or the test sensitivity rate from baseline information; and the final strategy involved the use of some combination of positive test results and observations (100 percent) and baseline information. All of these strategies involved mismanagement in the selection or combination of the given data needed to estimate correctly the diagnostic probability of the patient’s having the condition; none of these strategies led to the correct Bayesian estimate.

Table 2 (page 663) is a detailed breakdown of the diagnostic strategies used by students and
faculty members in the two vignettes. We listed percentages separately for students and faculty members for each vignette and as a combined total, weighted for sample size. The most commonly used strategy was to begin with the test sensitivity (that is, the likelihood that the test results will be positive for patients who, in fact, have the condition) and make either absolute or proportional adjustments based on information about the test sensitivity, resulting in false-positive observations. These represented 20 and 18 percent of the overall strategies.

In the next most likely strategy (16 percent), participants used only the test sensitivity and made no adjustments. Ten percent of the respondents adjusted the absolute positive observation (100 percent) by the chance of a false-positive observation. Twenty-six percent of students and faculty members gave some consideration to the baseline prevalence of the condition in reaching their diagnosis. Sixty-one percent of respondents included test sensitivity as part of their strategy. Fourteen percent of respondents used the test result itself (that is, they confused a positive observation with a positive condition) as an anchor for their diagnoses.

**Test sensitivity.** We found differences between students and faculty members with respect to the strategies used to determine diagnostic probability. These differences, apparent in both the 90 and 21 percent vignettes, are consistent with our hypothesis of greater diagnostic accuracy for faculty members. For example, faculty members were less likely to rely entirely on test sensitivity information. They also were less likely to make a relative value adjustment to the test sensitivity information for false-positive observations but were more apt to make an absolute adjustment. These differences were statistically significant and contributed to the overall significant difference (63 percent for students and 55 percent for faculty members) in respondents’ dependence on test sensitivity in some fashion ($P = .024$).

**Baseline information.** With respect to the use of baseline information, faculty members were significantly more likely than students to combine baseline information with test sensitivity data or test results and to use baseline information in some way (36 percent for faculty members versus 21 percent for students).

**Comments.** The comments offered by students and faculty members provided insight into the strategies they used to estimate diagnostic probability and how the respondents viewed this challenge. Neither group of respondents appeared to be comfortable with the task of making a precise estimate of diagnostic probability. Some comments reflect respondents’ frustration with working out a satisfactory answer: “There are too many things to consider.” “We have to leave room for error.” “I would ask the patient.” “I tend to favor aggressive treatments.” “There are too many variables to place a lot of faith in any conclusion.” “I would need to do the tests myself in order to know what to think.” “I need to know the risk factor first.” “You didn’t tell me what the sample size is.” “The sample size is too small.” “It’s going to depend on what the outcome of treatment is.” “This case is not specific enough.” “If I had any faith in research, I would want to keep my possibilities open.” “Most of these data seem irrelevant.” “Don’t ask me. I just sort of combine things.” Additional comments appeared to question the meaningfulness of making precise estimates of diagnostic probability. For example, “I really doubt that research can be truly accurate for supporting clinical decisions.” “We can only say that something is possible or not; we cannot assign numbers to it.” “It is foolish to extrapolate population studies to a single individual patient.” “This is an unrealistic case because I have no patients like this in my practice.” “Legal considerations have not been taken into account.” “It is better to be intuitive.” “These are hypotheses; nothing has been proven.” “There has been insufficient research in...
this area. “No P values have been provided.” “I would not rely on any data/research when considering the patient’s needs and treatment.” “I have seen several of these cases firsthand.”

<table>
<thead>
<tr>
<th>STRATEGY</th>
<th>OVERALL USE OF STRATEGY (%)</th>
<th>90% VIGNETTE</th>
<th>21% VIGNETTE</th>
<th>Combined Vignettes</th>
<th>90% VIGNETTE/21% VIGNETTE P VALUE*†</th>
<th>STUDENT/FACULTY P VALUE*‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students (n = 54)</td>
<td>Faculty members (n = 30)</td>
<td>Students (n = 56)</td>
<td>Faculty members (n = 20)</td>
<td>Students</td>
<td>Faculty members</td>
<td></td>
</tr>
<tr>
<td>Used baseline data only</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>11</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>Used test sensitivity data only</td>
<td>16</td>
<td>19</td>
<td>9</td>
<td>23</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>Used positive test result only (100%)</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Used false-positive observation only</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>Adjusted sensitivity for absolute value of false-positive observation</td>
<td>20</td>
<td>19</td>
<td>27</td>
<td>13</td>
<td>30</td>
<td>15</td>
</tr>
<tr>
<td>Adjusted sensitivity for relative value of false-positive observation</td>
<td>18</td>
<td>13</td>
<td>5</td>
<td>34</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>Adjusted positive observation (100%) for false-positive observation</td>
<td>10</td>
<td>21</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Adjusted baseline data for false-positive observation</td>
<td>8</td>
<td>13</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Used baseline data and test sensitivity in various relationships</td>
<td>6</td>
<td>6</td>
<td>18</td>
<td>0</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Used test sensitivity combined with baseline data</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Used positive observation (100%) combined with baseline data</td>
<td>5</td>
<td>4</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

**Combined Strategies**

<table>
<thead>
<tr>
<th></th>
<th>26</th>
<th>23</th>
<th>36</th>
<th>20</th>
<th>35</th>
<th>21</th>
<th>36</th>
<th>—</th>
<th>.038</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consideration of baseline data</td>
<td>61</td>
<td>56</td>
<td>59</td>
<td>70</td>
<td>50</td>
<td>63</td>
<td>55</td>
<td>—</td>
<td>.024</td>
</tr>
<tr>
<td>Consideration of test sensitivity</td>
<td>14</td>
<td>21</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>15</td>
<td>12</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

**Respondents’ Concerns**

<table>
<thead>
<tr>
<th></th>
<th>17</th>
<th>11</th>
<th>27</th>
<th>7</th>
<th>35</th>
<th>9</th>
<th>31</th>
<th>—</th>
<th>.001</th>
</tr>
</thead>
</table>
| Resistance to problem as given | * Only P values less than .05 are reported. Others are designated by a dash.† P values are for test of differences in proportions between the vignettes, with Bayesian analysis estimates of 90% or 21%; data are combined for students and faculty members.‡ Student/faculty P values are for test of differences in proportions between students and faculty members across vignette type.
These comments expressing frustration with the challenge of making precise estimates of diagnostic probability or expressing doubt about the value of using scientific data to make diagnostic estimates were three times as likely to come from faculty members than from students (31 versus 9 percent, \( P < .001 \)).

**DISCUSSION**

The results of this study confirmed our three hypotheses that were based on similar studies in medicine,\(^5,6,12-16,18-26,\)
- we observed a wide range of diagnostic probability estimates;
- when making estimates, participants gave undue weight to test evidence compared with baseline information;
- experienced practitioners were less prone to exaggerate the importance of test evidence than were students.

In addition, experienced practitioners expressed greater resistance to using structured approaches when estimating diagnostic probabilities, which suggests that the diagnostic process may involve other considerations, as noted below. Finally, these findings raise some issues regarding evidence-based dentistry.

**First hypothesis.** Standard deviations and interquartile ranges were large in this study, although generally smaller than those reported in the literature involving physicians’ diagnostic patterns.\(^4,12,13,15-21,24,25\) It is difficult to explain why students and practitioners, when given identical data, provided probability estimates that fluctuated so widely.

**Second hypothesis.** The second hypothesis—overemphasis on test outcomes—also was confirmed. For both vignettes and both groups of participants, the mean estimated diagnostic probability was between the baseline prevalence and the test sensitivity value, but it was closer to the test sensitivity value, a finding consistent with the literature.\(^5,6,12,13,15,20,21\) The fact that one-quarter of the participants gave no consideration to baseline data when calculating the diagnostic probability supports this hypothesis to an extreme degree.

**Third hypothesis.** Participants’ failure to make appropriate adjustments for baseline data is especially noteworthy. Faculty members were 1.7 times more likely than students to consider baseline information in their diagnoses, confirming our third hypothesis. Their mean estimated diagnostic probability was significantly closer to the Bayesian answer in both vignettes compared with that for students. This finding is consistent with those of other studies.\(^24,25\)

We might speculate that clinical practice builds a repertoire of baseline experience or at least an increasing awareness of the value of patients’ background information as an element of diagnosis. We do not need to assume that experienced practitioners consciously weighted background information in their calculations; the adjustments may have been largely intuitive. The finding by Aberegg and colleagues\(^28\) that experienced practitioners opt for more conservative or middle-of-the-road diagnoses is inconsistent with our results, because the baseline values in our study were more extreme than were the test sensitivity values in both vignettes.

An unexpected result of this study was that respondents, especially the experienced practitioners, exhibited a resistance to the structure required to estimate diagnostic probabilities precisely. Not only did participants complain that the task was difficult, but they challenged the concept of estimating precise probabilities as representing a best guess for their diagnoses. It is easy to form the impression that dentists do not have formulas for estimating the likelihood that a disease is present given various data that they use regularly. Respondents’ comments in this study seem to indicate that they were being asked to perform a mental process in which they did not regularly engage.

Since the mid-1960s, a line of research in applied decision theory (called “man as intuitive statistician”)\(^29\) has shown that we are capable of and typically use approximations of the more detailed rules used by decision scientists.\(^30,32\) In the aggregate, the results of our study confirmed that students and practitioners did use approximations of the best decision rules. However, the comments volunteered by participants suggest that something more is involved in the decision-making process.

Colombotos\(^33\) advanced a theory that health care professionals prefer to make personal, subjective assessments of clinical situations rather than data-driven, well-defined assessments. If it were possible for others with the same input information to calculate accurate diagnostic assessments, what would be left to professional judgment and the special place of practitioners? Furthermore, if others knew in some precise way
what was happening or what should happen with patients, then practitioners’ freedom of choice regarding treatment options would be truncated and others would be able to evaluate whether the clinician had performed correctly in a given situation.

In the Colombotos hypothesis, there can be too much diagnostic precision, or at least a desirable range of vagueness in which professional judgment holds sway. Ambiguity controlled by the health care professional has value. Man-Son-Hing and colleagues\(^34\) found that patients were more likely than practitioners to prefer exact quantitative diagnostic information. Berner and Maisiak\(^35\) reported that physicians resist using decision support systems that require an understanding of the system’s logic. Physicians also reported greater levels of discomfort with treatment regimens that limited their range of choices.\(^36,37\) Researchers also have identified the strategic advantage of using imprecise estimates of probabilities in business settings.\(^38-41\)

The research cited in this article assumes that dental care is based on diagnostic acumen and that an essential part of the diagnostic process is combining baseline information with evidence to generate an estimate of the likelihood that a patient has a specific disease or trauma condition. However, it is possible that this model is idealistic and more typical of researchers, while practicing dentists actually view the relationship between information, diagnosis and treatment paths differently.

Roswarski and Murray\(^42\) used vignettes to determine prescribing patterns among primary and emergency care physicians. Simple patient scenarios produced little decision bias; as the complexity increased, however, the amount of bias increased to the point at which some respondents deferred making choices altogether. Redelmeier and Shafir\(^19\) reported similar findings with several groups that evaluated a vignette involving osteoarthritis; they noted that complex cases led to respondents’ overreactions to the data (that is, salient evidence hypothesis) or maintenance of the status quo (that is, hypothesis of disengaging from the problem). Aberegg and colleagues\(^25\) investigated the diagnostic decision-making process of pulmonologists by using complex cases, and they found both status quo bias and a tendency of participants to undervalue information that would have led to treatment (both are forms of disengaging from the problem).

**Evidence-based dentistry.** The results of our research have implications for the discussion of evidence-based dentistry, which the American Dental Association defines as “an approach to oral health care that requires the judicious integration of systematic assessments of clinically relevant scientific evidence, relating to the patient’s oral and medical condition and history, with the dentist’s clinical expertise and the patient’s treatment needs and preferences.”\(^43\) This study appears to be the first to address the process of “judicious integration” in the context of dentistry. We also explored the effects of information with respect to the patient’s condition and the history of patients in the target population.

If the findings of this research are confirmed, our level of concern must be raised with regard to the possibility that salient, precise external evidence will receive undue weight in the decisions made by practitioners in individual cases. A report of Canadian primary care physicians concluded, “Clinicians strongly identified with the [evidence-based medicine] EBM model of clinical practice are less sensitive to context, which might be an obstacle to efforts to integrate patient values and clinical circumstances into patient-centered care.”\(^44(p1106)\)

Webster and colleagues\(^45\) found that primary care physicians generally ignore clinical guidelines because they perceive them to undervalue specific clinical situations. In a review of 76 studies, Cabana and colleagues\(^46\) found that the potential for harming individual patients in the course of following general rules was the principal reason cited by physicians who eschewed evidence-based medicine. If, in fact, many practitioners are worried about evidence-based dentistry, not because of the poor quality of the evidence, but because of concern that evidence could overshadow the particular dentist-patient relationship, then calls for greater methodological rigor through systematic reviews and meta-analyses may not be the most useful step forward.

**CONCLUSION**

The results of this study raise some issues regarding both the Bayesian model of diagnosis and evidence-based dentistry as the best conceptualizations of clinical practice. More research is needed to begin to understand how practitioners actually diagnose diseases rather than confirming the fact that they poorly approximate the way researchers advise them to diagnose disease. ■