

# “Is 28% Good or Bad?” Evaluability and Preference Reversals in Health Care Decisions

Brian J. Zikmund-Fisher, PhD, Angela Fagerlin, PhD, Peter A. Ubel, MD

*Choices of health care providers can become inconsistent when people lack sufficient context to assess the value of available information. In a series of surveys, general population samples were randomized to read descriptions of either 2 possible health care providers or a single provider. Some information about providers was easy to consider (e.g., travel time), but some was difficult to interpret without additional context (e.g., success rates). Ratings of the described health care providers varied significantly by whether options were evaluated independently or concurrently. For example, one*

*fertility clinic (33% success rate, 15 min away) was rated higher than a 2nd (40% success rate, 45 min away) when each clinic was considered separately (7.1 v. 6.2,  $P = 0.046$ ), but preferences reversed in joint evaluation (5.9 v. 6.7,  $P = 0.051$ ). The results suggest that clinicians and developers of patient information materials alike should consider information evaluability when deciding how to present health care options to patients. **Key words:** decision making; choice behavior; patient preferences; cognitive biases. (*Med Decis Making* 2004;24:142–148)*

As patients become more empowered health care decision makers, they are increasingly taking a more active role in the selection of their medical care providers. In addition to choosing the doctor or practice they will use for primary care, patients are also participating in the selection of specialty practices, such as obstetricians, and even the medical centers to be used for major treatments.

Patients have many types of information available for them to use when making such decisions. Some information is simple and straightforward, such as the location of a practice or the hours of operation, whereas other information may be more difficult for a layperson to evaluate. Lacking specific medical knowledge and expertise, patients may have little idea of how to interpret statistics such as the number of procedures performed by a surgeon or the pregnancy rate of an infertility clinic. Such information is often hard to evaluate and hence difficult to use in any decision process. As a result, patients might ignore hard-to-evaluate, but more important, information and instead choose a health care provider based on more trivial considerations.

The relevance of information evaluability to individual decision making has been demonstrated in the selection of both consumer products<sup>1–4</sup> and relative payoffs.<sup>5</sup> For example, Hsee<sup>1</sup> had individuals assess how much they would be willing to pay for multidisc compact disc players that differed in the number of CDs they held and in their sound quality as measured

Received 26 June 2003 from the Veterans' Affairs Health Services Research and Development Center for Practice Management and Outcomes Research, Veterans' Affairs Ann Arbor Healthcare System, Ann Arbor, Michigan (BJZ-F, AF, PAU); Program for Improving Health Care Decisions, Ann Arbor, Michigan (BJZ-F, AF, PAU); and the Division of General Internal Medicine, University of Michigan, Ann Arbor (AF, PAU). This work was presented as an oral presentation at the 25th annual meeting of the Society for Medical Decision Making, 22 October 2003. Financial support for this study was provided by a grant from the National Institutes for Health (R01 CA87595-01A1). In addition, Dr. Zikmund-Fisher is supported by an HSR&D postdoctoral fellowship from the Department of Veterans Affairs, and Dr. Ubel is the recipient of a Presidential Early Career Award for Scientists and Engineers. The funding agreement ensured the authors' independence in designing the study, interpreting the data, and publishing the report. The authors would like to acknowledge Cheralyn Boruvka, Andrew Englehart, Christine Goldstein, Molly Miklosovic, Jennifer Ream, and Todd Roberts for their research assistance and Tricia Eminger and 2 anonymous reviewers for comments on earlier versions. Revision accepted for publication 8 December 2003.

Address correspondence and reprint requests to Brian J. Zikmund-Fisher, Program for Improving Health Care Decisions, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-0429; e-mail: bzikmund@umich.edu.

DOI: 10.1177/0272989X04263154

by harmonic distortion percentages. Most nonaudiophiles do not know whether a total harmonic distortion of 0.01% is good or bad, but they do know that a 20-disc CD player could be more useful and valuable than one that holds only 5 CDs. As a result, when presented with a particular CD player by itself, most people were willing to pay the most for the players that hold the most CDs, regardless of their distortion ratings.

When people were presented with a pair of CD players, however, they could not only compare capacity but also assess the relative quality of the 2 players on harmonic distortion. A buyer may not know whether a 0.01% distortion is good or bad, but he or she does know that it is a lot more than 0.003% distortion. The joint presentation provides a context for the distortion values, making them more evaluable. As a result, when asked to give their willingness to pay for each of 2 CD players, one that holds 20 CDs and has distortion of 0.01% and the other with a capacity of 5 CDs but 0.003% distortion, people now gave higher willingness-to-pay values for the player with less distortion than they did for the player with a larger capacity, a significant preference reversal.

This preference reversal effect is not limited only to psychological experiments. Although the research in this area has generally used hypothetical scenarios presented to survey participants, the same type of "more-is-less" preference reversals have been demonstrated by economists using field experiments in the real-world market for collectable trading cards.<sup>6</sup>

Although information evaluability has been studied in consumer choice, we are unaware of any attempts to consider its role in medical decision making. This is an unfortunate oversight because health care decision making often hinges on information that is difficult to evaluate. For example, evaluability effects could lead to preference reversals when patients are trying to select a new doctor or hospital because information about medical care providers is often difficult for people without a medical background to understand.

## METHODS

### Questionnaire Design

To see whether evaluability differences lead to preference reversals in patients' selection of medical care providers, we presented 7 scenarios to a general population sample. Survey materials were directly adapted from Hsee's product selection tasks. Survey participants read a vignette describing a need for a new specialist physician or practice and then were randomly assigned to receive descriptions of either 2 possible

**Table 1** A Joint Evaluation Scenario

*Imagine that you have decided to have laser eye surgery. You contact some of the local eye surgery clinics and ask for information about the doctors who perform the surgeries.*

The 1st doctor you hear back from is Dr. Bettereyes. Dr. Bettereyes was educated at Harvard Medical School and uses a next-generation excimer laser that he purchased last year. He reports that he has performed this type of eye surgery about 80 times, with generally excellent results.

The 2nd doctor you hear back from is Dr. Seebetter. Dr. Seebetter was educated at the University of Iowa. His clinic uses the latest type of excimer laser, purchased 1 year ago. He reports that he has performed this type of eye surgery about 300 times, with generally excellent results.

*How good of a choice would Dr. Bettereyes be?*

candidates together or just 1 candidate. Specialists had 1 attribute that was hard to evaluate in isolation but easy to compare between alternatives (e.g., success rates or number of prior procedures completed) and 1 or more easy-to-evaluate features such as a patient testimonial or the distance to the physician's office. The primary dependent variable was participants' ratings as to how good of a choice each specialist would be to provide the required care.

Table 1 shows a sample scenario used in this study. In this example, a patient is seeking a physician to perform laser eye surgery. Approximately one-third of participants receiving this vignette read descriptions of both Dr. Bettereyes and Dr. Seebetter; one-third saw Dr. Bettereyes only, and one-third saw Dr. Seebetter only.

Both of the possible eye surgeons have an easy-to-evaluate characteristic: where they were educated. The fact that Dr. Bettereyes received his training at Harvard makes him an attractive candidate even in the absence of any other surgeons with whom to compare him. The

descriptions of both potential surgeons, however, include 1 harder-to-evaluate characteristic: his experience with the particular procedure under consideration. Does the fact that Dr. Bettereyes has performed this surgery 80 times represent a lot of experience or relatively little? A prospective patient, lacking specific medical knowledge of how often this procedure is performed, could find it difficult to determine the value of Dr. Bettereyes's experience without another physician to compare him to.

We hypothesized that, when considering the possible physicians separately, people would give higher ratings to Dr. Bettereyes on the strength of his Harvard education. The hard-to-evaluate information about experience would tend to receive relatively little decision weight in such separate evaluations. We predicted that only when individuals were presented with information about both physicians would the relative difference in experience become easier to evaluate and hence receive greater weight in the people's evaluations. A sufficiently large change in relative decision weight would result in the selection of Dr. Seebetter over Dr. Bettereyes in joint evaluation, a preference reversal across presentation methods.

## Subjects

Survey participants were recruited at the main cafeteria of a university hospital and at a university student center. Respondents received a packet containing several scenarios, sometimes by themselves and sometimes with additional survey materials for unrelated studies. These additional materials should not affect the results, however, since they involved a completely different topic and since the design for this study was fully randomized. Participants read each survey vignette and were then asked to rate each of the described medical care providers on a 0 to 10 scale according to "how good of a choice would [name of provider] be to [description of needed procedure]." These ratings comprised the primary dependent variable analyzed.

As the goal of this study was to determine the effects of difficult-to-evaluate information, we were concerned that health care providers might have context-specific knowledge that would help them to evaluate the information provided in the scenarios. To limit such problems, we therefore excluded data collected from anyone who described himself or herself as a health care practitioner (e.g., MD, DO, CRNP, or RN). This resulted in a total of 648 individuals who participated in the study. Of these, 58% were women, and the

average age of participants was 39 years (range = 18–90). The sample was predominantly white (85%) and well educated (45% with a bachelor's degree or higher), which is representative of the university community in which the study was conducted.

## Analysis

All analyses were performed using STATA 7.<sup>7</sup> We report *t* statistics rather than direct tests of the hypothesized interaction because the separate evaluation results compare 2 different participant groups (using unpaired *t* tests) while the joint evaluation data compare 2 judgments by the same individuals (i.e., paired observations).

## RESULTS

The 7 scenarios presented tested for evaluability biases using a variety of hard-to-evaluate information: success rates, complication rates, physician experience with a procedure, and the size of the medical program (i.e., institutional experience with a procedure). As shown by the attractiveness ratings presented in Table 2, preference reversals were found involving each of these characteristics. Survey participants consistently gave higher ratings in joint evaluation to the doctor or clinic with more experience, greater success rates, or fewer complications. Respondents reversed their preferences, however, when the 2 available alternatives were presented separately, demonstrating that the hard-to-evaluate attribute was de-emphasized in their decision making because of the lack of context. Tests of these changes in the reported attractiveness ratings were all significant.

For example, in the eye surgery scenario described above, the Harvard-educated Dr. Bettereyes was rated 6.4 out of a possible 10 when evaluated by himself, a slightly higher rating than the 5.9 received by the University of Iowa-educated Dr. Seebetter. However, when survey participants were given the opportunity to review both doctors and saw that Dr. Seebetter had almost 4 times as much specific experience as Dr. Bettereyes, they rated Dr. Seebetter a much better choice than Dr. Bettereyes (8.2 v. 5.7,  $t_{\text{joint}} = -5.28$ ,  $P < 0.001$ ), a highly significant shift in preference. A similar result is found in comparisons of obstetrics practices. When possible practices are evaluated one at a time, a nearby (10 min away) obstetrics practice with an 18% caesarian section rate receives significantly better attractiveness ratings than another practice that

**Table 2** Quality Ratings of Hypothetical Physicians or Medical Groups with Both Hard-to-Evaluate and Easy-to-Evaluate Attributes

	Survey Version			Preference Reversal
	Separate 1	Separate 2	Joint	
Eye surgeon				
<i>n</i>	58	53	49	
80 prior procedures, Harvard MD	6.4	—	5.7	
300 prior procedures, Iowa MD	—	5.9	8.2	
	$t_{\text{separate}} = 1.11$		$t_{\text{joint}} = -5.28^{**}$	Yes
Obstetrics practice				
<i>n</i>	55	54	47	
18% caesarian rate, 10 min away	7.4	—	6.9	
10% caesarian rate, 35 min away	—	6.3	7.3	
	$t_{\text{separate}} = 2.57^*$		$t_{\text{joint}} = -0.81$	Yes
Fertility clinic				
<i>n</i>	58	47	51	
33% success rate, 15 min away	7.1	—	5.9	
40% success rate, 45 min away	—	6.2	6.7	
	$t_{\text{separate}} = 2.02^*$		$t_{\text{joint}} = -2.00^{\dagger}$	Yes
Fertility clinic (testimonial version)				
<i>n</i>	55	58	59	
28% success rate, patient testimonial	6.3	—	6.1	
35% success rate	—	5.6	6.8	
	$t_{\text{separate}} = 1.97^{\dagger}$		$t_{\text{joint}} = -2.48^*$	Yes
Bronchial lavage				
<i>n</i>	73	71	72	
20–25 prior procedures, practices at major medical center, medical school faculty, physician testimonial	7.5	—	5.9	
100–120 prior procedures, practices at local community hospital	—	7.2	8.0	
	$t_{\text{separate}} = 0.86$		$t_{\text{joint}} = -6.19^{**}$	Yes
Bone marrow transplant center				
<i>n</i>	69	73	73	
45 transplants/year, family testimonial	7.4	—	6.4	
130 transplants/year	—	7.3	7.2	
	$t_{\text{separate}} = 0.29$		$t_{\text{joint}} = -2.33^*$	Yes
Primary care physician				
<i>n</i>	61	57	62	
Satisfaction rating 6.8/10, testimonial	6.4	—	6.0	
Satisfaction rating 7.6/10	—	6.2	6.8	
	$t_{\text{separate}} = 0.57$		$t_{\text{joint}} = -2.47^*$	Yes

Note: Ratings are on a 0 to 10 scale. Question format was “how good of a choice would [name of doctor or clinic] be to [description of procedure]?” The hard-to-evaluate attribute is listed first for each option.

<sup>†</sup> $P < 0.1$ . \* $P < 0.05$ . \*\* $P < 0.001$ .

is 35 min away but has only a 10% caesarian rate (7.4 v. 6.3,  $t_{\text{separate}} = 2.57, P < 0.05$ ). A side-by-side presentation, however, allowed participants to recognize the difference in complication rates, leading the more distant

practice to receive a higher rating (7.3) than the nearby one did (6.9 v. 7.3,  $t_{\text{joint}} = -0.81, ns$ ).

While the types of hard-to-evaluate information tested in these scenarios are all numerical, many differ-

ent kinds of easy-to-assess characteristics can influence patient choices sufficiently to create preference reversals. The scenarios presented used both numerical attributes that do not require context to be fully appraised, such as distance to the clinic or doctor's office, and categorical characteristics, such as a physician's medical school or hospital affiliations. We also demonstrated in 4 scenarios that the persuasive nature of a single patient's personal testimonial can sometimes be sufficient to overshadow hard-to-evaluate information in separate evaluations and create evaluation method preference reversals. In these scenarios, the weaker of the 2 alternatives on the hard-to-evaluate attribute was accompanied by a short patient testimonial that was written to be compelling without providing any substantive new information about the particular doctor or clinic. (See the appendix for the text of the patient testimonials.)

For example, 1 scenario considered 2 possible bone marrow transplant centers for a patient with acute myelogenous leukemia. In this scenario, the "Linden Center" performs approximately 3 times as many transplants as does the "Stewart Center" (130 v. 45). However, when each center is evaluated independently, it is difficult to know whether the 45 transplants performed each year by the Stewart Center implies that it is a large or a small program. Furthermore, the description of the Stewart Center is accompanied by a testimonial by a family member of a transplant survivor from that program. Although the testimonial provides no new information about the program beyond the fact that his sister survived a transplant performed there, the combination of his positive testimonial and the difficulty of assessing program size data causes the Stewart Center to be very slightly preferred in separate evaluation (7.4 v. 7.3). The difference in program size, however, clearly is perceived as important when comparing both centers side-by-side, leading the Stewart Center to be perceived as significantly inferior in joint presentation (6.4 v. 7.2,  $t_{\text{joint}} = -2.33$ ,  $P < 0.05$ ).

## DISCUSSION

### Why Evaluability Matters

The scenarios used in this study examined the types of information generally available to patients when choosing their health care providers. The results demonstrate that provider selection decisions can change depending on whether the patient is considering a physician in isolation or whether 2 or more providers are being compared side by side.

These findings are of concern because the high ratings people gave to certain physicians in separate evaluation could be an indication that they would pick this doctor or clinic without looking at any other choices. If patients suspend their search for other health care providers early on, they run a greater risk of selecting a physician they would not have chosen if they had known the full set of alternatives. In that case, both care and satisfaction will likely suffer.

Although the present inquiry considered only provider choice decisions, the same psychological processes likely affect treatment decisions whenever patients lack the experience or medical knowledge to fully understand the meaning of the statistics presented to them. In particular, both treatment success rates and the percentage of patients experiencing side effects are difficult to evaluate in the absence of comparative information. Thus, patients may be prone to choosing treatments that are minimally invasive or less costly (attributes that are easy to evaluate) because they cannot assess the relative value of a particular success or side effect statistic. By the same token, personal anecdotes ("my sister had that done") will influence patient decision making the most when quantitative comparisons of other attributes are difficult to perform. Only when all alternatives are presented concurrently will such hard-to-evaluate information be fully integrated into choice processes.

Evaluability concerns are also relevant to the design of patient information tools. Informational materials are often generated for complex treatment decisions for which the patient needs to consider many different issues, such as possible risks, advantages, side effects, and success probabilities. If the information to be included in the brochure is of a type unfamiliar to the target audience, it is likely to be comparatively ignored in patients' decision making unless additional context is provided.

There are multiple ways to improve the evaluability of unfamiliar information. Side-by-side comparison of treatment alternatives automatically provides some context (i.e., relative position), but information evaluability can be improved even in separate evaluation by telling patients additional facts about the attributes in question.<sup>4</sup> Providing average or "normal" values for an attribute improves evaluability by clarifying whether a value is high or low. Evaluability is further improved by telling patients the range or distribution of possible values on each attribute. So, for example, a prospective mother can effectively evaluate an obstetrics practice with a 26% caesarian section rate even in the absence of a direct comparison if she knows that the

average practice in her area has a 22% caesarian rate and local practices range from 10% to 31%.

### Which Evaluation Method Is Best?

The discussion above suggests that in many cases, side-by-side comparisons offer significant advantages over separate presentations of the available options. However, joint evaluation is not always preferable. Two scenarios that highlight this fact are the fertility treatment scenarios reported in Table 2. In both cases, participants compared 2 large fertility clinics (each performing 200 in vitro fertilization cycles per year) that differed in their success rates over the past 6 months. One clinic, "Fertility Partners," had a 7% lower success rate than its competitor, "Reproductive Associates," but was either located 30 min closer or described in a personal testimonial from a happy past patient. As shown in Table 2, in both fertility treatment scenarios, survey respondents rated Fertility Partners lower than Reproductive Associates in joint evaluation, but it received higher average ratings than its competitor when the clinics were evaluated one at a time.

At first glance, respondents' preference for Reproductive Associates when comparing the 2 options side by side seems appropriate. But is Reproductive Associates really the better clinic? The clinic description includes information about the sample size from which the success statistics were generated (approximately 100 patients over 6 months), and thus one can determine whether the observed difference in pregnancy rates is noteworthy. In fact, a proportions test of the difference in pregnancy rates is not even close to significant ( $z = 1.03$ ,  $P < 0.3$ ). Thus, respondents in the joint evaluation condition are expressing a strong preference for the Reproductive Associates clinic on the basis of a difference in success rates that quite possibly occurred by chance.

The results from the fertility treatment scenarios make clear the fact that the choice of evaluation method is situation dependent. Neither separate evaluation nor joint presentation is always appropriate. In most applied contexts, and certainly whenever alternatives have many unique attributes such as enthusiastic patient testimonials, joint evaluation will usually be the preferred method. Comparative presentations should, in principle, allow both easy- and hard-to-evaluate attributes to receive proportionate decision weight. As long as all the attributes are relatively equal in importance, joint evaluation facilitates the use of decision-making strategies that integrate all available information.

On the other hand, when numerical information may be receiving excessive weight, separate presentation of each alternative allows for greater consideration of other important issues. For example, prospective parents are prone to focus exclusively on a fertility clinic's success rates, which may be highly variable due to small sample sizes. As a result, they may essentially ignore other important attributes, such as proximity or treatment philosophy, which can contribute to both the couple's satisfaction and the likelihood of a successful outcome. Consistent with this idea, it is interesting to note that the Centers for Disease Control and Prevention's Assisted Reproductive Technology Reports,<sup>8</sup> which provide detailed statistics regarding the outcomes of in vitro fertilizations and other infertility treatments at certified clinics in the United States, discourage side-by-side comparison of individual clinics.

The choice of presentation format is therefore a function of the goal of the interaction. Clinicians and decision aid developers alike need to consciously consider how well their intended audience will be able to evaluate each type of information to be conveyed and select a presentation mode that facilitates the objectives of the communication.

## APPENDIX

### Patient Testimonials

#### Fertility Clinic

"I went to Fertility Partners for treatment of my infertility and conceived my daughter on my very first IVF cycle. It was such a relief. I thought the staff and doctors were very professional, and my daughter is proof that they know what they're doing."

#### Bronchial Lavage

"My patient had been having really severe shortness of breath, which was keeping him from his passion: golf. He was very frustrated because he couldn't even practice for long without getting winded, and a full round was out of the question. After Dr. Myers did the bronchial lavage, he couldn't

wait to tell me how much better he felt. He actually called me from the course the first time he played 18 holes again.”

### **Bone Marrow Transplant Center**

“The Stewart Center had a team of doctors and nurses trained to work with stem cell transplant patients. My sister was pretty weak from the cancer by the time she got to the center, and she needed blood transfusions every few days. The transplant was very difficult, but it was really her only

chance at beating the leukemia. Now she’s back to being herself again.”

### **Primary Care Physician**

“Dr. Feelbetter is just what I want in my doctor. He always asks about old problems to make sure they’re resolved and he helps me think about ways I can get exercise in spite of my schedule. I feel comfortable with him as my doctor.”

---

### **REFERENCES**

1. Hsee CK. The evaluability hypothesis: an explanation for preference reversals between joint and separate evaluations of alternatives. *Organ Behav Hum Decis Process*. 1996;67:247–257.
2. Hsee CK. Less is better: when low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making*. 1998;11:107–21.
3. Hsee CK, Leclerc F. Will products look more attractive when presented separately or together? *Journal of Consumer Research*. 1998;25:175–86.
4. Hsee CK, Blount S, Loewenstein GF, Bazerman MH. Preference reversals between joint and separate evaluations of options: a review and theoretical analysis. *Psychol Bull*. 1999;125:576–90.
5. Bazerman MH, Loewenstein GF, Blount White S. Reversal of preference in allocation decisions: judging an alternative versus choosing among alternatives. *Admin Sci Q*. 1992;37:220–40.
6. List JA. Preference reversals of a different kind: the “more is less” phenomenon. *Am Econ Rev*. 2002;92:1636–43.
7. Stata statistical software [computer program]. Version 7. College Station (TX): Stata; 2001.
8. Centers for Disease Control and Prevention. 2000 Assisted Reproductive Technology Success Rates: National Summary and Fertility Clinic Reports. Atlanta (GA): Department of Health and Human Services; 2002. Available from: <http://www.cdc.gov/nccdphp/drh/art.htm>