

What's Time Got to Do with It? Inattention to Duration in Interpretation of Survival Graphs

Brian J. Zikmund-Fisher,^{1,2,3*} Angela Fagerlin,^{1,2,3} and Peter A. Ubel^{1,2,3}

Reports of randomized clinical trials often use survival curves to summarize clinical outcomes over time and graphically demonstrate evidence of treatment effectiveness. Survival curves can also be used in patient communications to display how health risks accumulate over time. In a randomized survey experiment, administered online, we tested whether people viewing survival curves appropriately adjust their risk perceptions to account for the duration shown. Internet users ($N = 864$) were recruited from a demographically balanced U.S. panel. Participants read about a hypothetical disease and then viewed one of four survival graphs that displayed mortality risks with and without treatment. Survival graphs showed either a visually large or visually small difference between treatments and were labeled to represent either 5-year or 15-year risk statistics. Participants then provided ratings of disease seriousness, as well as treatment effectiveness for each possible treatment. Variations in ratings corresponded more with visual dissimilarity than with changes in the statistical risk exhibited, with participants perceiving somewhat greater disease seriousness and significant differences in treatment effectiveness in large visual difference graphs. We conclude that when people interpret survival curves, they often fail to sufficiently account for the timeframe represented and perceive more risk and larger differences when identical risks are displayed over longer periods of time. We recommend that all presentations of survival graphics, whether to patients, physicians, or scientists, emphasize duration information (e.g., in the title) and remind readers that attending to graph axis labels is the only way to pierce these visual illusions.

KEY WORDS: Cognitive biases; risk communication; visual displays

1. INTRODUCTION

Imagine that you are reading research reports describing two clinical trials that compare treatments for a life-threatening condition. In each case, the trial

data are reported and summarized in a survival curve graphic, here presented as Figs. 1 and 2. Each figure graphically represents the proportion of the patient population with this medical condition surviving over time if the patients were to take Pill A as treatment, Pill B as treatment, or No Pills. Now, before you read on, take a moment to ask yourself the following questions: In which population was the hypothetical Crawford's Disease more serious? Does Pill B seem more effective in Fig. 1 or Fig. 2? In which case does Pill A provide a greater therapeutic advantage when compared to Pill B?

The two survival graphs shown in Figs. 1 and 2 display identical annual mortality risks. In both cases, patients faced a 7.5% annual mortality risk if they

¹ VA Health Services Research & Development Center for Practice Management and Outcomes Research, VA Ann Arbor Healthcare System, Ann Arbor, MI, USA.

² Division of General Internal Medicine, University of Michigan, Ann Arbor, MI, USA.

³ Center for Behavioral and Decision Sciences in Medicine, Ann Arbor, MI, USA.

* Address correspondence to Brian J. Zikmund-Fisher, Center for Behavioral and Decision Sciences in Medicine, 300 North Ingalls Building, Rm. 7C27, Ann Arbor, MI 48109-0429, USA; tel: (734)-936-9179; fax: (734)-936-8944; bzikmund@umich.edu.

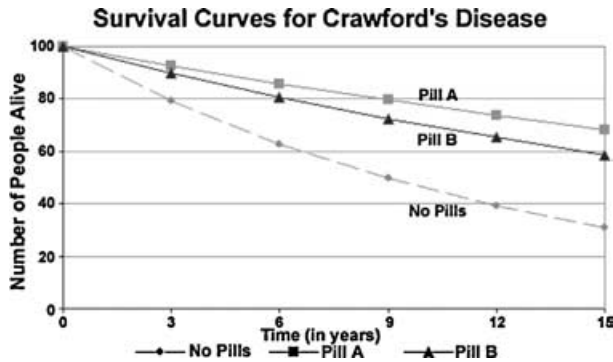


Fig. 1. Survival graph displaying hypothetical mortality risks (Graph B: 15-year medium risk condition).

receive no treatment, a risk that is reduced to 2.5% per year if they take Pill A or to 3.5% per year if they take Pill B. As these risks are compounded annually (i.e., in year T , 7.5% of those alive in year $T - 1$ die without treatment), all of the survival curves are exponential in shape. The two graphics differ only in the length of time displayed: Fig. 2 is simply Fig. 1 truncated at 5 years of data.

Yet, if you are like many people, the two survival graphics may not seem identical. This difference in perception is a result of a visual illusion that inhibits accurate risk perception (Lau & Ng, 2002). In a survival graph (in fact, in any line graph), both the slope of a curve and the area between two curves varies with the length of time displayed. The shorter the time period shown, the flatter the slopes and hence the smaller the area between two curves. Accurate perception of both overall disease risk and treatment effectiveness therefore requires a mental adjustment. An impression of low risk created by viewing

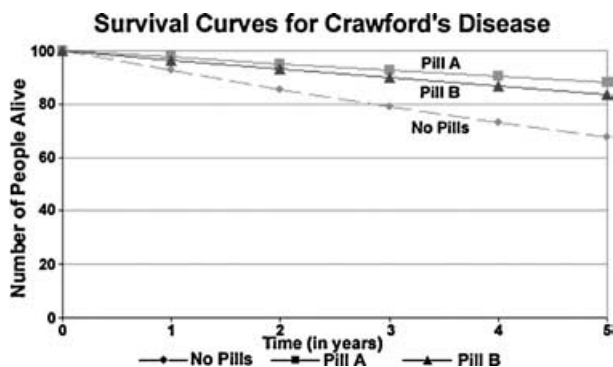


Fig. 2. Survival graph displaying hypothetical mortality risks (Graph C: 5-year medium risk condition).

relatively flat survival curves must be consciously increased if the graph presents data compiled over a short period of time, while perceptions of significant risk must be mediated and reduced if the graph displays many years of risk information. If people fail to make such adjustments, then changing the time period shown in a survival graph can result in under- or overappreciation of important differences in mortality risks.

The fact that visual slope underlies line graph perception is not new; Huff's (1954) *How to Lie with Statistics* discussed its implications 50 years ago, and the concept still receives major attention in current treatises on design of quantitative graphics (Tuft, 2001). The resulting visual illusions are of particular concern in medicine, however, because survival graphics are commonly used to communicate disease risk and treatment effectiveness information in reports of randomized clinical trials. (For example, see Fisher *et al.*, 2001; Landolfi *et al.*, 2004.) As the movement toward evidence-based medicine draws such reports into the everyday experience of primary care providers, clinicians will increasingly be expected to take their understandings of outcomes data (often reported in survival graphs) and incorporate that knowledge into their clinical practice.

Patients, too, are being introduced to survival graphs as decision aids and patient information materials attempt to communicate how risk accumulates over time, a key issue in discussions about chronic health conditions and risk-reducing interventions. Although comprehension of survival graphs by patients is not always consistent (Armstrong *et al.*, 2002; Mazur & Hickam, 1994; Mazur & Merz, 1993), research has demonstrated that patients can understand these curves and that use of such graphics can improve risk understanding compared to numerical presentations (Armstrong *et al.*, 2001; Mazur & Hickam, 1990, 1993, 1994, 1996; Mazur & Merz, 1994). To our knowledge, however, no research exists examining the consistency (or lack thereof) of risk perceptions derived from survival graphs that differ in the length of time displayed.

Our study tested whether risk perceptions obtained after viewing different survival graphs corresponded appropriately with variations in statistical risk. We hypothesized that the similarity or dissimilarity of the visual image presented in survival graphics would show greater correlation with risk perceptions than the displayed risk levels.

2. METHODS

2.1. Overview of Study Design

In an Internet-administered survey, participants read a description of a hypothetical health condition. We randomized participants to receive one of four survival curve graphics and then assessed individuals' perceptions of disease seriousness and treatment effectiveness.

2.2. Participants

Study participants were drawn from a panel of Internet users who voluntarily agreed to participate in research surveys. This panel is administered by Survey Sample International (SSI) and includes over 1 million unique member households recruited through random digit dialing, banner ads, and other "permission-based" techniques. (For more information, see http://www.surveysampling.com/ssi_home.html.) Individuals completing our web-based survey were entered into a drawing to win a cash prize of up to \$1,000. E-mail invitations were sent to a sample of panel members stratified to mirror the U.S. census population based on age, gender, education level, and income.

2.3. Intervention

Respondents read a brief description of a hypothetical condition called Crawford's Disease. Patients

with this condition have a constant mortality risk over time but can reduce that risk by taking one of two medications. (See Appendix A for the full text of the scenario.) Survey participants then viewed one of four graphs indicating the proportion of Crawford's Disease patients who took either Pill A, Pill B, or No Pills still surviving as time since diagnosis increases.

Fig. 3 summarizes the four experimental conditions in this study, which comprise a 2 × 2 factorial design. We created two types of survival graphs, "Large" visual difference graphs (A and B) and "Small" visual difference graphs (C and D). These graphics differed in the slopes of the survival curves (steeper vs. flatter) and hence also the size of the spaces between different curves (steeper curves create larger spaces between survival curves). Within each visual difference pair (A vs. B, C vs. D), the two graphs differed only in the labeling of the time period displayed: one version was labeled so that it displayed 5 years, of mortality statistics while the second was labeled as displaying 15 years of data. So, while Graph A shows large visual differences occurring over 5 years, Graph B (Fig. 1) has identical slopes and areas between the lines but is labeled as displaying 15 years of data. By the same token, Graph C (Fig. 2) displays a small visual difference graphic labeled to represent 5 years of risk information, while Graph D is the same graph relabeled to show 15 years of data.

The four graphs can also be categorized by the levels of statistical risk displayed. As noted in Table I,

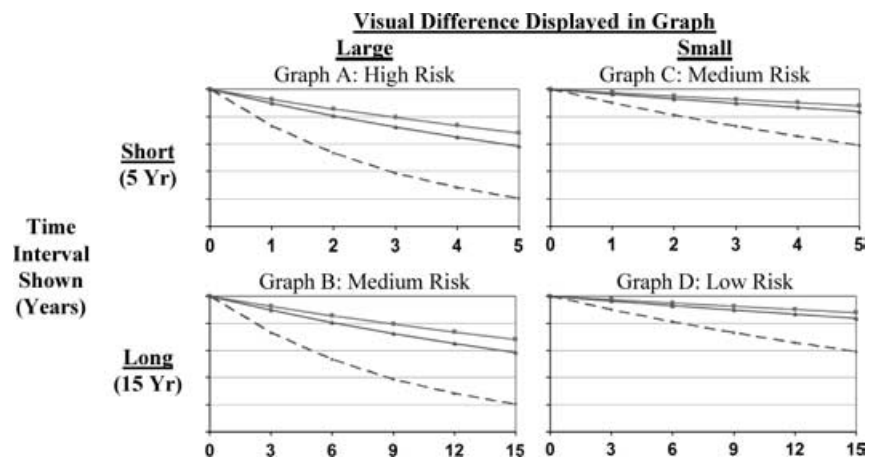


Fig. 3. Experimental conditions.

Notes: Graph details that were constant across conditions are omitted here; see Figs. 1 and 2 for fully detailed images of Graphs B and C, respectively. Horizontal lines represent 100, 80, 60, 40, 20, and 0 people alive, respectively. Visual difference pairs (Graphs A and B, Graphs C and D) display identical line slopes and areas between curves, except with different time (X) axis labels and hence different annual risk levels. See Table I for the corresponding annual risk levels.

Table I. Details of the Four Survival Graphs Used

Annual Mortality Risk With	Graph A	Graph B	Graph C	Graph D
	(%)	(%)	(%)	(%)
No Pills	20.9	7.5	7.5	2.6
Pill A	7.3	2.5	2.5	0.8
Pill B	10.1	3.5	3.5	1.2

Note: Risk was compounded annually, yielding exponential survival functions.

annual mortality risk was exactly identical in the two medium risk graphs, Graph B (large visual difference, 15-year duration) and Graph C (small visual difference, 5-year duration). Furthermore, since survey participants were explicitly told that “the chance of dying from Crawford’s Disease is the same every year no matter how long you live,” viewers of these two graphs received equivalent risk information, merely displayed for different periods of time. The two remaining graphs, however, display substantially different statistical risk levels. Graph A (high risk) displays the highest annual mortality risks, almost three times larger than those displayed in the two medium risk graphs (B and C), while Graph D (low risk) displays risks only about one-third as large as the medium risk graphs.

2.4. Outcome Measures

Our outcome measures were ratings of disease seriousness for Crawford’s Disease and treatment effectiveness for both Pill A and Pill B. All ratings were made on 11-point (0–10) scales, with endpoints labeled as “Not At All Serious” and “Extremely Serious” for disease risk assessments and “Not At All Effective” and “Extremely Effective” for treatment effectiveness ratings. In addition, we took the difference between the effectiveness ratings for Pill A and

Pill B as a measure of the perceived benefit of using one treatment versus the other. Because this measure is a within-subject difference, it controls for variations in how respondents interpreted the response scale and thus should have greater sensitivity.

2.5. Hypotheses

Table II summarizes how the four risk graphics compare with each other in terms of the statistical risk level reported, the resulting absolute and relative risk reductions of treatment (ARR and RRR), and the graphs’ visual similarity or dissimilarity. We hypothesized that perceptions of disease seriousness would primarily be generated by the visual slope of the survival curve rather than the true mathematical risk represented. We therefore predicted that ratings of disease seriousness would be equal for the pairs of graphs with identical visual images (A and B, C and D) but differ significantly between Graphs B and C, despite the fact that these two graphs display identical risk information. Since Graph B displays steeper survival curves, we expected respondents to perceive greater disease seriousness with Graph B than with Graph C. Regarding treatment effectiveness, the relative risk reduction (RRR) associated with Pill A or Pill B was the same in all four conditions. However, we predicted that perceptions of effectiveness would be driven by the size of the area between two curves. Thus, we hypothesized that participants would rate treatments as more effective, either absolutely or in comparison with each other, when viewing either of the large visual difference graphs (Graphs A and B) versus either of the small visual difference graphs (Graphs C and D).

2.6. Statistical Analysis

We utilized *t*-tests to determine whether respondents’ ratings differed across conditions, with results

	Survival Graph Comparisons			
	Graph A vs. Graph B	Graph B vs. Graph C	Graph C vs. Graph D	Graph A vs. Graph D
Annual disease mortality risk	≠	=	≠	≠
Treatment absolute risk reduction (ARR)	≠	=	≠	≠
Treatment relative risk reduction (RRR)	=	=	=	=
Visual differences	= ^a	≠ ^a	= ^a	= ^a

Table II. Comparing Statistical Risk and Visual Similarities across the Four Graphs

^aGraphs A and B display survival curves with steeper slopes and larger visual areas between curves than Graphs C and D.

Note: = Equal risk levels; ≠ Different risk levels.

confirmed using nonparametric Wilcoxon rank-sum (Mann-Whitney) tests. Two types of comparisons were performed. We compared ratings within each pair of visually similar graphs (Graphs A and B, C and D) to determine if participants recognized the significant differences in objective risk displayed. We also compared ratings between the two medium risk conditions (Graphs B and C), since those two graphics display identical annual mortality risks in visually different formats. All analyses were performed using STATA Version 8.

3. RESULTS

A total of 6,372 people received email messages inviting them to participate in an online survey, and 1,022 (a 16.0% response rate) clicked the embedded link to begin the survey. While some dropout occurred, 864 participants (84.5%) provided usable data. The mean age of our sample was 48 years (range 18–83 years), 37% were male, and, of the 702 who reported racial and ethnic background information, 92% described themselves as Caucasian, 4% as Hispanic, and 3% as African American. In addition, 15% of our sample rated their overall health as either “fair” or “poor.” There were no significant variations in sample characteristics between the groups viewing the four different survival graphics.

Table III summarizes the results of the statistical analyses comparing disease seriousness and treatment effectiveness ratings between the four graphics. Even though both of the 5-year graphs (Graphs A and C) displayed substantially higher mortality risks than their 15-year counterparts (Graphs B and D), we found no significant differences (all $ps > 0.1$) in

participants’ ratings of disease risk between graphs with similar visual images (A vs. B and C vs. D). Perceptions of treatment effectiveness were also equivalent across these pairs (all $ps > 0.1$), accurately reflecting the fact that the relative risk reduction (RRR) provided by each possible treatment was the same in all conditions.

The comparisons between Graphs B and C (the visually different graphs showing identical annual risks), however, showed some notable variations. The comparisons of disease seriousness ratings were marginally significant ($p = 0.095$), with higher disease seriousness observed with the 15-year duration Graph B. A larger effect emerges, however, when we subtract the two effectiveness ratings to see how much difference our respondents perceived between Pill A and Pill B and thus control for individual differences in response scale usage. The average difference in effectiveness ratings between treatments is 1.15 on our 11-point scale in the Graph C (small visual difference, 5-year) condition but increases to 1.61 in the Graph B (large visual difference, 15-year) condition. This difference is highly significant using either a t -test of the means ($t = -3.47$, $p < 0.001$) or a Wilcoxon rank-sum test of the response distributions ($p < 0.001$).

4. DISCUSSION

Our results indicate that people often fail to appropriately adjust the risk perceptions they derive from viewing survival curves to account for different lengths of time displayed. Visually similar survival graphs yield similar perceptions, regardless of how many years of data they are marked as displaying. In addition, our data suggest that extending the

Table III. Study Results

	Survival Graph Comparisons							
	Graph A ($n = 242$)	vs.	Graph B ($n = 189$)	vs.	Graph C ($n = 231$)	vs.	Graph D ($n = 202$)	
Results								
Disease seriousness ratings	7.87	= $p = 0.227$	7.61	≈ $p = 0.095$	7.24	= $p = 0.162$	6.92	
Pill A effectiveness ratings	7.40	= $p = 0.622$	7.49	= $p = 0.151$	7.21	= $p = 0.128$	7.50	
Pill B effectiveness ratings	5.80	= $p = 0.524$	5.91	= $p = 0.406$	6.07	= $p = 0.473$	6.20	
Difference between effectiveness ratings of Pill A and Pill B	1.61	= $p = 0.998$	1.61	≠ $p < 0.001$	1.15	= $p = 0.132$	1.32	

Notes: All comparisons performed using t -tests. ≠ Significantly different ratings ($p < 0.05$); ≈ Marginally different ratings ($p < 0.1$); = Not significantly different ratings ($p \geq 0.1$).

length of time displayed in a survival graph tends to increase both perceived disease seriousness and the degree to which differences in survival are viewed as significant. The latter finding is consistent with the foreground salience effect shown in previous research on risk communication graphics (Stone *et al.*, 2003).

Several factors limit the generalizability of our findings. First, while the sample invited to participate matched the U.S. population on several key demographic characteristics, we did observe some response biases that led our final sample to be more female and Caucasian than the U.S. population as a whole. Second, although the 16% response rate we achieved is typical of Internet studies, this figure is lower than is commonly expected for mailed surveys. Our randomized experimental design, however, controls for any observed or hypothesized response biases. Our goal was to demonstrate differences between experimental conditions, rather than to assess any particular characteristic of the population, and sample peculiarities such as the overrepresentation of women occurred evenly across the groups viewing the four different graphics. We did, however, lack the power to identify any racial or ethnic disparities in respondents' reactions to the risk data presented. (Such disparities have been demonstrated in global assessments of environmental and health risks (Finucane *et al.*, 2000; Palmer, 2003).) Third, the graphs studied here displayed risks that were constant over time; people may be more or less able to develop accurate perceptions of risk and treatment effectiveness when viewing survival curves showing risks that temporally vary, as occurs in most real-world settings. Finally, our use of Internet subjects and hypothetical scenarios may have reduced participants' motivation to make the cognitive effort required for accurate comprehension. It may be that doctors reading important clinical trial results and patients reviewing materials about their own pressing medical issues would be more accurate than our subjects were, although we doubt that interest alone would be sufficient to fully correct for these types of misinterpretations.

Still, the present study suggests that clinicians and medical researchers need to pay particular attention to the time dimension of survival graphs when interpreting clinical research reports. The duration of a clinical trial is determined by a variety of factors, both scientific and practical. Yet, our findings imply that the amount of follow-up data provided may change beliefs about treatment effectiveness *by itself*. By the same token, communications to patients about their long-term health risks (e.g., breast cancer risks or

diabetes complications) could be interpreted substantially differently depending on whether the decision aid developer chooses to present short-term (e.g., 5-year) risk data or long-term (e.g., lifetime) statistics.

Awareness of this cognitive bias both in the design of survival graphics and in their interpretation by readers may help to improve risk communications. We recommend that all presentations of risk graphics (survival or otherwise), whether to patients, physicians, or scientists, emphasize duration information by specifically identifying the time period displayed in the title (preferred) or legend. Instructions that highlight duration information, especially those that put the focal statistics into context, may also be helpful. (For example, "Remember, this is your risk of developing cancer *in the next 5 years*; your lifetime risk will be higher and depends on your age.") Publishers (academic or general) should also consider standardizing the spacing used to denote units of time in their graphics, so that 15-year graphs are three times *wider* than 5-year graphs are. In the end, however, the burden remains on the reader to focus his or her attention on the graph axis labels and consciously use that information to mediate the perceptions of treatment effectiveness derived from survival graphs.

ACKNOWLEDGMENTS

Financial support for this study was provided by grants from the National Institutes for Health (R01 CA87595-01A1 and 1 P50 CA101451-01). During this study, Dr. Zikmund-Fisher was supported by an HSR&D Post-Doctoral Fellowship from the U.S. Department of Veterans Affairs, and Dr. Ubel was the recipient of a Presidential Early Career Award for Scientists and Engineers (PECASE). The funding agreements ensured the authors' independence in designing the study, interpreting the data, and publishing the report. The authors would also like to acknowledge the web programming and research assistance provided by Todd Roberts and the technical support of MSInteractive.

APPENDIX: TEXT OF THE HYPOTHETICAL DISEASE SCENARIO

Page 1

Crawford's Disease

Imagine there is a rare, serious disease called Crawford's Disease. The chance of dying from this

disease is the same every year no matter how long you live.

However, doctors agree that people who have Crawford's Disease have a better chance of surviving if they take one of two pills every day for the rest of their lives. These are called Pill A and Pill B. The chance of dying from the disease when you take one of the pills is the also same every year no matter how long you live, but the chance is smaller than if neither pill is taken. The two pills cost about the same and are paid for by most insurance companies.

Pretend that 300 people were diagnosed with Crawford's Disease at year 0. A group of 100 people started taking Pill A, a second group of 100 people began taking Pill B, and a third group of 100 people did not receive any pills. On the next page, you will see information about what happens to each of these groups.

Page 2

This graph is called a *survival curve*. It shows the number of people with Crawford's Disease who survive each year for the first 5[15] years after being diagnosed with the disease. The chance of dying from Crawford's Disease is the same every year no matter how long you live.

The pink line with the squares stands for the people who took Pill A. The blue line with the triangles stands for the people who took Pill B. The dashed green line with circles represents the people who didn't take any pills. Each of the squares (Pill A), triangles (Pill B), and circles (No Pills) stand for the number of people who are alive at each point in time.

Page 3

How *serious* do you think Crawford's Disease is? (Select one)

Not at all serious
Extremely serious

Page 4

How *effective* do you think *Pill A* is for treating Crawford's Disease? (Select one)

Not at all effective
Extremely effective

How *effective* do you think *Pill B* is for treating Crawford's Disease? (Select one)

Not at all effective
Extremely effective

REFERENCES

- Armstrong, K., Fitzgerald, G., Schwartz, J. S., & Ubel, P. A. (2001). Using survival curve comparisons to inform patient decision making: Can a practice exercise improve understanding? *Journal of General Internal Medicine, 16*(7), 482–485.
- Armstrong, K., Schwartz, J. S., Fitzgerald, G., Putt, M., & Ubel, P. A. (2002). Effect of framing as gain versus loss on understanding and hypothetical treatment choices: Survival and mortality curves. *Medical Decision Making, 22*(1), 76–83.
- Finucane, M. L., Slovic, P., Mertz, C. K., Flynn, J., & Satterfield, T. A. (2000). Gender, race, and perceived risk: The "white male" effect. *Health, Risk & Society, 2*(2), 159–172.
- Fisher, B., Dignam, J., Bryant, J., & Wolmark, N. (2001). Five versus more than five years of Tamoxifen for lymph node-negative breast cancer: Updated findings from the national surgical adjuvant breast and bowel project B-14 randomized trial. *Journal of the National Cancer Institute, 93*(9), 684–690.
- Huff, D. (1954). *How to Lie with Statistics*. New York: W.W. Norton.
- Landolfi, R., Marchioli, R., Kutti, J., Gisslinger, H., Tognoni, G., Patrono, C., & Barbui, T. (2004). Efficacy and safety of low-dose aspirin in polycythemia vera. *New England Journal of Medicine, 350*(2), 114–124.
- Lau, E. W., & Ng, G. A. (2002). Visual illusions created by survival curves and the need to avoid potential misinterpretation. *Medical Decision Making, 22*, 238–244.
- Mazur, D. J., & Hickam, D. H. (1990). Interpretation of graphic data by patients in a general medicine clinic. *Journal of General Internal Medicine, 5*, 402–405.
- Mazur, D. J., & Hickam, D. H. (1993). Patients' and physicians' interpretations of graphic data displays. *Medical Decision Making, 13*(1), 59–63.
- Mazur, D. J., & Hickam, D. H. (1994). The effect of physician's explanations on patients' treatment preferences: Five-year survival data. *Medical Decision Making, 14*(3), 255–258.
- Mazur, D. J., & Hickam, D. H. (1996). Five-year survival curves: How much data are enough for patient-physician decision making in general surgery? *European Journal of Surgery, 162*, 101–104.
- Mazur, D. J., & Merz, J. F. (1993). How the manner of presentation of data influences older patients in determining their treatment preferences. *Journal of the American Geriatrics Society, 41*(3), 223–228.
- Mazur, D. J., & Merz, J. F. (1994). Five-year survival data in surgical decision making: What aspects of graphical data influence patients' preferences? *Theoretical Surgery, 9*, 76–81.
- Palmer, C. G. S. (2003). Risk perception: Another look at the "white male" effect. *Health, Risk & Society, 5*(1), 71–83.
- Stone, E. R., Sieck, W. R., Bull, B. E., Yates, J. F., Parks, S. C., & Rush, C. J. (2003). Foreground: Background salience: Explaining the effects of graphical displays on risk avoidance. *Organizational Behavior and Human Decision Processes, 90*, 19–36.
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

