Reviewer's report

Title: Risk Communication Formats for Low Probability Events: A study of patient preferences

Version: 1 Date: 11 October 2007

Reviewer: Brian Zikmund-Fisher

Reviewer's report:

General
This manuscript presents the results of a study that compared 6 graphical approaches to communicate the effectiveness of two cancer screening options. The study used a convenience sample of 29 patients and employed an Analytic Hierarchy Process to identify relative preferences. The displays themselves are innovative in several ways: 1) a bar graph format and a grouped icon format that each had secondary sub-graphs that magnified the relevant information about the low risk %'s (<6%) shown, 2) a flow chart format that uses natural frequencies to communicate the risk information, and 3) three combination presentations that combined pairs of the other formats. The authors argue that their results support three conclusions regarding risk communication of small (<5-6%) risks: a) that graphs that use magnified views of key areas are acceptable to patients, b) that natural flow diagrams may be effective, and (c) that patients prefer multiple graph formats.

While the literature is in desperate need of studies comparing different approaches to risk communication in varying contexts and the authors' overall methodology appears reasonably sound, the stimulus images used here differ in many dimensions, so that a clear attribution of the observed preference differences cannot be made to the choice of format. In addition, the data collected do not appear sufficient to cleanly support two of the three conclusions stated.

The larger message I would like to convey is that in studies of risk graphics, the devil really is in the details. If colors, labeling, text use, and other dimensions are maximally comparable across graph types, then effects such as preference differences or comprehension deficits can be fairly attributed to the choice of graph formats. Such studies are extremely valuable. Unfortunately, this study does not provide such clear findings.

-------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1) The most significant concern is that the study design aims to compare different graphs, yet the graphs that are used differ in many aspects other than the format.
While the risk statistic information is the same across the different graph types (and hence one might be able to compare people's ability to read the graphs), the output measure here was restricted to _preferences_. Thus, any variations between the stimuli that are not related to format are confounds to the results. An abbreviated list of such issues follows:

---Bar chart: Included numerical axis labels and uses colors with clear contrast vs. background. But, format breaks up possible outcomes into three pairs of bars in a way that makes it visually difficult to realize that the sum of each of the three bars of the same color represents the full range of outcomes (as opposed to stacked bars which do show the denominator). Groupings of outcomes here are by possible outcome (thus, the two bars representing # of cancers are next to each other). Numerical text information shown in each bar facilitates comprehension of exact numbers.

---Grouped icon display: has no axis labels to denote fractional amounts or guide gist perceptions (in contrast to bar), uses different core colors to represent each outcome than were used in the bar graph (thus potentially conveying different information, since colors have affective meanings), has a low unit to background area ratio (reducing signal to noise), and has a colored background that reduces contrast (making it harder to visually see what proportion of area represents each outcome). Icon displays by definition group outcomes by screening option (so all three possible outcomes for option A are together, dissimilar to bar format). No numerical text information is provided anywhere in the image, thus making it more difficult to discern exact percentages and further creating informational differences vs. the bar format.

---Flow diagram: This fundamentally different format has the least amount of non-data visual elements but uses yet a third set of colors to denote options. Numerical information is provided.

Because of the high number of differences between graph formats, the resultant differences in preference cannot be attributed to the format design alone or even primarily. The fact that participants preferred multiple graph formats is similarly difficult to attribute. Because different formats differed in the types and number of visual cues provided, providing multiple graphs at once may have created complementarities -- one graph's strengths may have offset the deficits of the other. Whether multiple formats would still be preferred when graphs of equal informational value were paired remains unclear.

2) The conclusion "that bar charts and icon displays that include magnified views of differences between decision alternatives are acceptable to patients" appears only marginally supportable from this data. Clearly, participants expressed _relative_ preferences for certain graphs vs. others. But, no control graphs (e.g., simpler bar or icon graphs WITHOUT magnified views) were provided, and thus we cannot know whether participants would have chosen or rejected the study stimuli when presented with simpler alternatives. The authors frame their conclusions in an absolute sense, i.e. that these graphs are in some sense
sufficiently acceptable for potential future use, but it would seem that some type of absolute judgment of acceptability is needed. If the only advance is that it is now shown that this type of graph can be created, why did we need the participants?

3) The conclusion "that natural frequency flow diagrams may be as effective as other currently recommended risk presentation formats" is also unsupportable with this data. Effectiveness implies accurate comprehension and ability to be used by readers to accomplish larger tasks. The authors elicitations of preference do not support judgments of effectiveness, and they do not report any other measures (e.g., comprehension, use in identifying optimal treatment choices, etc.) that might. While participants did appear to express some degree of preference for this format, language on at the top of p.10 should be modified to reflect the finding of preference and not effectiveness.

4) p4. "Small risks are difficult to display in bar charts and icon displays." Not really, it's just that small risks mean that the colored area is relatively small when compared to the total. In fact, differences between two small risks may be easier to discern than equal changes between two larger risks, because the 0 point provides a more relevant standard of reference. Greater discussion and clarity regarding why investigating "small" risks is needed. The traditional uses for "magnifiers" in risk communication have been for situations where the risk is smaller than 1%, a level of detail that neither bar nor icon displays are designed to handle effectively. Clarification of why risks in the 2%-5% range are particularly different from risks in the 20+% range is needed to justify the significance of the study's focus on this type of image.

5) The first page and a half of the discussion does not discuss the results at all but simply summarizes existing research (and would be hence more relevant to the introduction). A clearer summary of the results, with direct connections of the results to literature (akin to the discussion of the flow diagrams on p 9) would tighten the flow of the paper.

6) p.8: there's no need to report a p value smaller than 0.001. It's significant, and using more decimal points comes across as almost obnoxious.

7) Why did the authors include both bar and pie charts to visually represent the preference dimension? This seems to create a very visually complicated task, especially since it involves showing additional graph types at the same time as the subjects are trying to evaluate other graph types.

8) I have other concerns about the fact that preferences were elicited as a trinary
response and then strength of preference was elicited separately. This method seems likely to lead to more extreme responses than if a single scale were used to represent preference for 1 presentation vs. the other. A justification would be helpful.

**What next?:** Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests