

Medical Decision Aides: An Evaluation Methodology for Information Design

Debra Satterfield, Sunghyun R. Kang, Paul Bruski, Brytton Bjorngaard, Fred Malven* and
Nora Ladjahasan**

* Iowa State University, Department of Art and Design
USA, debra815@iastate.edu, Shrkang@iastate.edu, bruski@iastate.edu,
malven@iastate.edu, bryttonb@iastate.edu

** Iowa State University, Institute for Design Research and Outreach, USA
nading@iastate.edu

Abstract: On a daily basis, many people are expected to make important medical decisions for themselves or for those in their care. Some decisions are routine, however other decisions have potentially life-altering consequences. Therefore, the design of the information tools, referred to as decision aids (DAs), used by medical providers must meet the social, emotional, and cognitive expectations of the physician, as well as the patient, in order to be truly effective and useful in the clinical setting. This research is the first step in creating a body of information in the area of evidence-based information design.

This information design methodology combines evaluation techniques or strategies from anthropology, usability, education, and unique methods developed specifically from the research of medical information. The development process of this methodology will be discussed from the concept development stage, to focus groups, and through the final data collection process.

Key words: *Medical decision aids, Information design, Evaluation methodology*

1. Introduction

Medical decision aid (DAs) tools are designed to help patients understand their options and consider the personal preferences of benefits and harm in making specific and deliberative choices [1, 2]. According to the Cochrane Collaboration review, *Decision Aids for People Facing Health Treatment or Screening Decisions*, the use of decision aids compared to usual healthcare measures help people feel more comfortable with their decisions. This is made evident, they say, through reduction in scores on decision conflict scales. People who use decision aids generally feel more informed about treatment options and clearer about their decisions with regard to personal values [3]. Evidence suggests that when patients are actively involved in the treatment decision-making process, they are likely to suffer less over it and achieve better performance in their treatments [4].

DAs can take the form of printed materials, web sites, videotapes, or audiotapes. According to Coulter and Entwistle [5], there were no preferences in the type of media delivery; however there were concerns about the information content. This research showed that people preferred structured and concise materials that included pictures and diagrams.

1.1 Medical Decision-Making Guidelines

Significant research and publication on medical decision aids has been conducted by the Ottawa Hospital Research Institute (OHRI) at the Ottawa Hospital and the University of Ottawa. The Patient Decision Aid research group at Ottawa was founded in 1995 to help patients and practitioners make tough healthcare decisions. This organization as part of the Ottawa Health Decision Centre has set policies for the design and evaluation of DAs. The Development Toolkit is comprised of a conceptual framework, development methods, evaluations measures, international standards, the Cochrane Systematic Review document, and the Cochrane Decision Aid Registry [6].

The Ottawa Decision Support Framework takes into consideration the interrelated elements of decisional needs, decision quality and decision support. Decision aids come under the “decision support component” whose role is to clarify decisions and needs, provide facts and probabilities, clarify values, guide/coach/support skills, and monitor/facilitate progress.

However, even with these guidelines, the quality of decision aids varies greatly and a debate exists about what should be the underlying concepts for decision aids and a lack of an agreed upon criteria for the development of these decision-making tools. And, because DAs can have a profound impact on patient decision-making, it is important that developers use recognized methods, avoid bias, and cite valid evidence sources [7].

1.2 Visual Guidelines for Medical DAs

While some guidelines exist for the informational content of decision aids, no such guidelines exist with regard to the visual information design of DAs. In addition, no research has been conducted to identify the relationships between the design of visual information and the information content of the DA. Currently, the only mention of any visual information design criteria comes under development methods section of the Ottawa Development Toolkit in the International Patient Decision Aid Standards (IPDAS) section II, which says, “Development Process: Does the Patient Decision Aid Present information in a balanced manner? Shows negative/positive features with equal detail (fonts, order, display of statistics) 9.2. [6]” And it is questionable whether the intention of this standard is in reference to the design of the visual information or is only intended to be a guideline for visual clarity or readability.

Because no criteria for the visual design of DAs has been developed and systematically tested for validity, this research seeks to develop a method for both evaluating and designing visual information for DAs. The criteria that must be considered for the visual design of DAs includes typographic variables, fonts, color, symbols, layout, positive/negative space, photos, visual hierarchy, stylistic treatments, material usage, print or electronic format, motion graphics, sound, and other visual, physical, and sensory elements of communication. This research, which focuses on a holistic evaluation of design variables in the context of an existing medical DA, is the first step in creating a body of information in the area of evidence-based visual information design. It is the intention that this research will produce evaluation and design guidelines based on user input and perceptions to create a more appropriate user experience for patients and practitioners who use DAs for making complex medical decisions. Because of the nature of medical decision making, it was determined that the social and emotional aspects of information design must be evaluated in addition to the usability and readability aspects of the designs.

2. Methodology

This methodology evaluates social and emotional factors as they apply to the medical decision-making process. Factors such as the social stigma attached to certain medical conditions such as epilepsy or sexually transmitted diseases such as HIV were discussed with regard to their impact on decision-making. Because medical decisions are often times made in consultation with close friends or family members, this collaborative process was considered with regard to its impact on how medical decision aides should be designed and evaluated. This information design methodology combines evaluation techniques or strategies from anthropology, usability, education, and unique methods developed specifically for this type of medical information.

An initial list of social and emotional words associated with feelings and attitudes was generated. These words were then put into pairs of opposing concepts such as hot vs. cold. Using focus group input, this list of words (see Table 1) was then evaluated with regard to how people perceived them in the context of evaluating medical DAs. Based on feedback from these focus groups, the words overwhelming, eye catching, and unappealing were removed from the final word list.

The remaining opposing words were subjected to Pearson correlation to determine if there is a negative correlation between the paired words. Ten out of 14 pairs of words turned out to have negative correlation with the opposite word (i.e. the more healthy the DA is the less sad it is).

Table 1. Word Pairs

Word 1	Word 2	Sig. (2-tailed)	Pearson correlation
Healthy	Sickly	.010	-.916*
Trustful	Deceitful	.010	-.915*
Interesting	Boring	.001	-.974*
Clear	Ambiguous	.030	-.855*
Relevant	Useless	.011	-.913*
Simple	Complicated	.004	-.948**
Efficient	Time Consuming	.203	-.605
Empowering	Manipulative	.864	-.091
Functional	Useless	.001	-.980**
Personal	Generic	.060	-.793
Educational	Instructional	.186	-.624
Comforting	Frightening	.011	-.914*
Happy	Sad	.002	-.968**
Comprehensive	Limited	.000	-.984**

* correlation is significant at the 0.05 level (2-tailed)

** correlation is significant at the .01 level (2-tailed)

A medical decision aid designed for selecting a treatment protocol for diabetes was selected to be the DA evaluated during the development phase of the methodology. This DA consists of a set of six cards. Each card focused on an issue associated with a life style or medical aspect of diabetes. They were analyzed with regard to their visual design based on the elements of color codification, title, typography, symbol usage, quantity of text, and the visual density of information (See Figure 1).

The six DA cards were designed using a set of consistent visual parameters. The 5"x9 1/4" cards each have a white background with a 1" top bar. The top bars each have a different color and a one or two line title written in either black or white text. The text on the cards is all set in Franklin Gothic with bold used for the headlines, the five treatment categories, and for emphasis on certain words in the body copy of card 6. The title text is in 36 point type and sub-heads in 18 point type. The body copy was in 14 point or below in a normal

weight. The cards are laminated with a heavy plastic that is matte on the front and gloss on the back of the card. The corners are cut in a rounded shape. The top of the body copy is consistently aligned 3/8" below the color bar, however, the side and bottom margins varied between cards.



Figure 1. Visual Density

2.1 Focus Groups Study

Four focus groups were conducted to help evaluate the appropriateness of the word pairs and to observe people's behaviors, actions, and spoken words when evaluating the DAs. The focus groups were recruited by word of mouth and the first participant recruited was encouraged to recruit one or two familiar people so that the final group would be a social group of friends or family members. Because medical decisions are often made in consultation with close friends or family, it was felt that the focus groups should reflect this social dynamic. The focus groups represented a variety of age, gender, and cultural groups. The focus group participants were encouraged to interact with the DAs and discuss their feeling about them. The focus group discussions were video taped and the tapes later analyzed with regard to the opinions and reactions expressed by the focus group participants with regard to the DAs and the word pairs. The findings from the focus groups were used to inform the final the survey questions. Any words that caused confusion were eliminated and additional words or concepts were gleaned from the dialogue of these focus groups.

2.2 Test design

The test was designed to use all six DAs from the diabetes decision aid. The survey was designed to rank the six DAs on a scale of 1 to 6, with one being the least associated with a word and 6 being the most associated with a word, for each of twenty-eight words from the word pair list. The word pairs were tested separately to evaluate if they would actually be perceived as opposing concepts using an independent ranking system. The test was administered using the online survey tool, Survey Monkey. However, the survey was designed to allow the participant to hold, shuffle, and interact with the DA while a test administrator entered the responses into a computer. However, the final demographic portion of the test was designed to have the participant enter the information into the computer independently in order to avoid any anxiety to the survey participant based on revealing personal information to the test facilitator. The test was designed to take approximately 20 – 30 minutes and subjects were not compensated for their participation.

2.3 Testing process

Survey participants were recruited using posters, e-mail recruitment, and word of mouth advertising methods. The tests were conducted over an approximately five-month period and 107 participants were recruited. Between survey questions, the participants were encouraged to pick up and randomize the DAs before proceeding to the next word. When the word was given, the test facilitator instructed the survey participant to order the DAs from least to most with regard to the word being assessed. The test facilitator would then enter the assigned value for each DA into the web-based survey. At the conclusion of the word assessments, the participants were given control of the computer and allowed to enter their own demographic information. At the conclusion of the survey, the test facilitator encouraged the survey participant to offer any information that they wanted to give about their impressions of the DAs or their rationale for any responses that they gave during the survey. This qualitative data was digitally recorded and used to create three value-based scales for analysis.

2.3 Data analysis

The cards were evaluated based on their survey results and their visual design assessment. They had associated mean values for each of 28 words from the survey (See Table 2). These values were then tested with regard to their validity based on their correlation with their score on the opposing word from the survey. Of the 14 pairs of words, 10 pairs were found to have negative correlation with their corresponding opposing words.

The significant positive words were then subjected to factor analysis to determine if they measure the same concept. Reliability testing was done to validate the scales created.

Table 2. Mean Value

	Card 1	Card 2	Card 3	Card 4	Card 5	Card 6
Healthy	2.73	2.67	4.83	4.43	3.54	2.86
Sickly	3.66	4.29	2.72	3.14	3.44	3.71
Trustful	2.64	2.6	3.83	4.48	3.03	4.45
Deceitful	3.91	4.57	2.8	2.89	3.94	2.9
Interesting	1.85	3.53	4.68	4.21	3.87	2.86
Boring	4.8	3.44	2.34	3.25	3.05	2.34
Clear	3.35	1.94	4.35	3.95	3.44	3.96
Ambiguous	3.98	4.77	3.17	2.98	3.82	2.31
Relevant	2.65	3.09	4.38	4.42	3.04	3.43
Useless	4.51	4.35	2.51	2.7	3.96	2.91
Simple	4.33	2.07	4.75	3.46	3.44	2.79
Complicated	2.49	5.19	2.45	3.6	3.68	3.61
Efficient	3.71	2.34	4.57	4.23	3.45	2.73
Time Consuming	2.2	4.43	2.82	4.01	3.27	4.27
Empowering	2.27	2.8	4.73	4.59	3.36	3.27
Manipulative	3.08	4.06	3.23	3.5	3.95	3.17
Functional	2.4	2.43	4.98	4.64	3.08	3.87
Useless	4.51	4.35	2.51	2.7	3.96	2.91
Personal	2.1	2.68	4.84	3.98	3.96	3.44
Generic	4.95	3.22	2.89	3.08	3.03	3.82
Educational	2.07	2.81	3.57	4.15	3.28	5.12
Instructional	2.1	2.87	4.87	4.75	2.87	3.64
Comforting	2.68	2.4	4.99	4.46	3.6	2.88
Frightening	3.51	4.58	2.22	3.05	3.5	4.14

Happy	2.72	3.21	5.46	4.12	3.68	1.83
Sad	3.68	3.99	2.09	3.11	3.5	4.65
Comprehensive	1.84	2.89	3.89	4.77	3.09	4.54
Limited	5.29	3.76	3.11	2.47	3.84	2.44

The three scales were created from this data (See Table 3). The first scale, emotion, is based on survey data collected from the words healthy, interesting, relevant, comfortable, and happy. The second scale, cognition, is based on data from the words trust, functional, and comprehensive. The third scale, function, is based on the words clear and simple. The scales were created by adding the data from these words and dividing by the number of words associated with the scale. This provides a convenient way of comparing the values.

Table 3. Scales Mean Value

Card #	Emotion	Cognition	Function
1	2.52	2.29	3.84
2	2.98	2.64	2
3	4.87	4.23	4.55
4	4.33	4.63	3.71
5	3.54	3.07	3.44
6	2.77	4.29	3.38
Survey Words	healthy	trust	clear
	interesting	functional	simple
	relevant	comprehensive	
	comfortable		
	happy		
Reliability test	0.955	0.945	0.806

Reliability is the consistency of the measurement, or the degree to which an instrument measures the same way each time it is used under the same condition with the same subjects (See Table 3). In short, it is the repeatability of the measurement. A measure is considered reliable if a person's score on the same test given twice is similar. It is important to remember that reliability is not measured, it is estimated. One common way of computing correlation values among the questions on your instruments is by using Cronbach's Alpha. As a general rule, reliability should not be below .80 for widely used scales. [8]

3. Findings and Discussion

3.1 Card One

Of the six cards, Card 1 has the lowest percentage of visual information on it. The title bar is a deep red color with the words "Blood Sugar" and, below in a second line and in a smaller point size, the words, "(A1c Reduction)." Both lines of text are written in white. The information on the card uses no symbols and only 13 words of drug names and associated percentages. The printed content utilizes only 15% of the space on the card. The five treatment categories are in the same weight and point size, however the numerical data (percentages) are inconsistent in size and weight. This creates some visual discord; however the overall appearance of the layout is clean and minimal. The use of red in the title bar may influence how people view this card based on feelings with regard to blood or blood draws.

From the survey, Card 1 received six strongly negative associations from the tested words (see table 3). It received mean values of 4 or greater for the words deceitful, boring, ambiguous, useless, simple, generic, and limited. Among the six cards, it was the most highly associated with the words boring, useless, generic, and limited.

Based on the three scales derived from the survey data, Card 1 had a low associated value with emotion, cognition, and function. It was slightly below average on the emotion and cognition scales which is consistent with its survey results and its visual design assessments, both of which indicate this card is limited in its appeal and its quantity of information. The functional scale is slightly above the average mark which is consistent with the limited but clear quality of its information.

3.2 Card 2

Of the six cards, Card 2 has the highest percentage of visual information on it. The title bar is a deep gold color with the words “Low Blood Sugar” and, below in a second line and in a smaller point size, the word, “Hypoglycemia.” Both lines of text are written in black. The information on the card uses a combination of grey, yellow, and red dots as symbols to visually represent percentages associated with risk. There are a total of 35 words of information on the card. The printed content utilizes only 54% of the space. The five treatment categories are in the same weight and point size, however the symbolic data in the form of dots is dramatically different from the treatment categories in weight. This creates some visual discord; however the overall appearance of the layout is highly compartmentalized and structured.

From the survey, Card 2 received nine strongly negative associations from the words tested. It received mean values of 4 or greater for the words sickly, deceitful, ambiguous, useless, complicated, time consuming, manipulative, frightening, and sad. Among the six cards, it was the most highly associated with the words sickly, deceitful, ambiguous, complicated, time consuming, and frightening.

Based on the three scales derived from the survey data, Card 2 also had a low associated value with emotion, cognition, and function. It was slightly below average on the emotion and cognition scales which is consistent with its survey results and its visual design assessments, both of which indicate this card is difficult to understand and seems complicated. The functional scale even lower than the other two and this may reflect that fact that many of the people surveyed just didn't like the format and didn't understand the data. Interestingly, when actually scrutinized by this research team, the total number of dots, which are supposed to visually suggest a percentage of risk, did not even total 100. Thus making the percentage reference invalid. This further supports the skepticism indicated by the survey respondents.

3.3 Card 3

Card 3 has a moderate quantity of visual information on it. The title bar is a yellow green color with the words “Daily Routine” written in black. The information on the card uses a combination of text for the treatment categories and symbols for food, medicine, and injections. There are a total of 20 words of information on the card. The printed content utilizes only 25% of the space. The five treatment categories are in the same weight and point size, however some categories have more text or symbols than others. This creates some imbalance between the treatment categories; however the overall appearance of the layout is visually engaging because of the colored icons and structured due to the text and visual delimiters.

From the survey, Card 3 received thirteen strongly positive associations from the words tested. It received mean values of 4 or greater for the words healthy, interesting, clear, relevant, simple, efficient, functional, personal, instructional, comforting, happy, and manageable. Among the six cards, it was the most highly associated with the word instructional.

Based on the three scales derived from the survey data, Card 3 scored high with regard to emotion, cognition, and function. All three scales received scores higher than 4 indicating a strong positive association with emotion, cognition, and function. Of the six cards, Card 3 scored the highest on both the scale of emotion and function. This is consistent with the survey data and reinforces the fact that survey participants saw this card as visually interesting and easy to read or interpret. The cognitive score is still reasonably high and reinforces the idea that the information is perceived as sufficient.

3.4 Card 4

Card 4 has a large quantity of visual information on it. The title bar is a dark green color with the words “Daily Sugar Testing” and, below in a second line and in a smaller point size, the word, “(Monitoring).” Both lines of text are written in white. The information on the card uses a combination of text for the treatment categories and a chart with red dots to indicate monitoring frequency and a 2-4 line block of explanatory text. There are a total of 56 words of information on the card. The printed content utilizes only 32% of the space. The five treatment categories are in the same weight and point size, however some of the monitoring frequency charts have more red dots than others. The overall appearance of the layout is somewhat visually engaging because of the charts and red dots. This card appears structured due to the justification of the text and charts and it appears compartmentalized due to the use of visual delimiters.

From the survey, Card 4 received sixteen strongly positive associations from the words tested. It received mean values of 4 or greater for the words healthy, trustful, interesting, clear, relevant, efficient, time consuming, empowering, functional, personal, educational, instructional, comforting, happy, comprehensive, and manageable. However, among the six cards, it was not the most highly associated card for any of the words.

Based on the three scales derived from the survey data, Card 4 also scored high with regard to emotion, cognition, and function. The scales for emotion and cognition both received scores higher than 4. In addition, the scale of function is above a 3 indicating that this is still in a reasonable range. Of the six cards, Card 4 scored the highest on the cognition scale. This is consistent with the survey data and reinforces the fact that survey participants saw this card as having good quality information. The high emotion scale score also indicates that the information is not perceived as overwhelming or uncomfortable. The functional score is slightly lower probably due to the larger quantity of information on this card and its varying form.

3.5 Card 5

Card 5 has a moderate quantity of visual information on it. The title bar is a dark purple color with the words “Weight Change” written in white. The information on the card uses a combination of text for the treatment categories and a horizontal bar that is comprised of closely positioned grey boxes with white plus or minus signs arranged going out from a neutral center point. Starting in the center, the boxes are colored in a range from light blue to magenta to indicate weight gain and in a range from light blue to dark blue to show weight loss. There are a total of 28 words of information on the card. The printed content utilizes only 29% of the space. The five

treatment categories are in the same weight and point size. The overall appearance of the layout is somewhat visually engaging due to the color transitions in the horizontal bars. This card appears very compartmentalized due to the use of the horizontal bars as visual delimiters.

From the survey, Card 5 it was not the most highly associated card for any of the words among the six cards. Based on the three scales derived from the survey data, Card 5 received average scores with regard to emotion, cognition, and function. This is consistent with the survey data and reinforces the fact that survey participants did not see this card as a prime example of any of these three areas.

3.6 Card 6

Card 6 has the highest quantity of visual information on it. The title bar is a dark magenta color with the words “Side Affects” written in white. The information on the card uses only text for the treatment categories and for the explanation of the side affects. In addition, certain words such as nausea, indigestion, diarrhea, fluid retention (edema), and rash have been somewhat randomly emphasized in the text through the use of a bold weight text. There are a total of 131 words of information on the card. The printed content utilizes only 44% of the space. The five treatment categories are in the same weight and point size. The overall appearance of the layout is complicated due to the relatively large quantity of text. This card appears very full and text-heavy.

From the survey, Card 6 received the most varied results from the words tested. It received mean values of 4 or greater for the words trustful, clear, time consuming, educational, frightening, sad, and comprehensive. Among the six cards, it was the most highly associated with the word sad.

Based on the three scales derived from the survey data, Card 6 scored high with regard to cognition and function. However, its score on emotion was below average. This is consistent with the survey data and reinforces the fact that survey participants saw this card as clear, functional, and comprehensive. But the lack of visual images or icons kept this card from scoring higher in the areas of emotion. While it seemed factual, it did not seem interesting to most survey participants.

4. Conclusions

The survey results and the scales developed from the survey data in this research have direct relationships to design variables. The three scales developed in this study will be used for future analysis to identify relationships based on various demographic information such as gender, income levels, language skills, field of study and education levels to determine what the exact relationship is with regard to a scale rating and a visual design variable.

Areas for future research are the demographic groups that demonstrated statistically significant differences. These included gender, age, income level, and native language. The combination of quantitative and qualitative data used in this research will be used to determine the exact interpretation of these findings and apply it to a design methodology for developing the visual design and testing of future medical DAs.

5. Acknowledgment

In collaboration with Dr. Victor Montori, Maggie Breslin, and the Knowledge Encounter Research unit at the Mayo Clinic, we developed a method for testing the impact of specific design variables used in medical decision aids as they are perceived by various demographic groups. This important on-going research will be used to

improve the quality of medical discussions held during the clinical encounter between patients and medical providers.

This research is supported by Iowa State University through the Bailey Research Career Development Grant. We would like to thank our undergraduate research assistants, Becky Murphy, Melissa Dilling, and Gina Assmann, for their hard work and support in conducting the surveys.

6. References

- [1] Graham, I. D. et al (2007) Physician's intentions and use of three patient decision aids, *BMC Medical Informatics and Decision Making* [Online]. Available at <<http://www.biomedcentral.com/1472-6947/7/20>> [Accessed 30 April 2009]
- [2] Nelson, W., and Han, P., Fagerlin, A., Stefanek, M., and Ubel, P. (2007) Rethinking the Objectives of Decision Aids: A Call for Conceptual Clarity, *Medical Decision Making*, Vol. 27, No. 5, pp 609-618.
- [3] O'Connor, A.M., Bennett, C.L., Stacey, D., Barry M., Col N.F., Eden K.B., Entwistle, V.A., Fiset, V., Holmes-Rovner, M., Khangura, S., Llewellyn-Thomas, H. and Rovner, D. Decision Aids for People Facing Health Treatment or Screening Decisions. *Cochrane Database of Systematic Reviews* 2003, Issue 1. Art. No.: CD001431. DOI: 10.1002/14651858. CD001431. p 45.
- [4] Coulter, A., and Entwistle, V. (1999) Sharing decisions with patients: is the information good enough?, *BMJ* 318, pp 318-322.
- [5] *ibid*
- [6] Ottawa Decision Support Framework [Online]. Available at <<http://decisionaid.ohri.ca/odsf.html>> [Accessed 13 May 2009]
- [7] Elwyn, G., O'Connor, A., Stacey, D., Volk, R., Edwards, A., Coulter, A., Thomson, R., Barratt, A., Barry, M., Bernstein, S., Butow, P., Clarke, A., Entwistle, V., Feldman-Stewart, D., Holmes-Rovner, M., Llewellyn-Thomas, H., Mounjid, N., Mulley, A., Ruland, C., Sepucha, K., Sykes, A. and Whelan, T. on behalf of the International Patient Decision Aids Standards (IPDAS) Collaboration (2006), Developing a quality criteria framework for patient decision aids: online international Delphi consensus process, *BMJ*, doi:10.1136/bmj.38926.629329.AE, p 1.
- [8] Carmines, E. G., and Zeller, R.A.(1990) *Reliability and Validity Assessment*, Sage University Paper, New Bury Park, p5.