

Getting Tools Used

Lessons for Health Care from Successful Consumer Decision Aids



U.S. News & World Report:
AMERICA'S BEST COLLEGES



CONSUMER REPORTS:
CAR BUYING GUIDE



eBay.com



NUTRITION FACTS PANELS

NOTE: THIS IS ONLY A PORTION OF THE GETTING TOOLS USED RESEARCH REPORT. FOR THE FULL DOCUMENT AND OTHER INFORMATION VISIT WWW.CFAH.ORG.

Table of Contents

| | |
|--|-----|
| Foreword by Jessie Gruman..... | 1 |
| Executive Summary..... | 3 |
| About CFAH..... | 7 |
| Table of Contents..... | 8 |
| Introduction: 21 st Century Marketplace..... | 9 |
| Research Framework..... | 17 |
| Case Studies..... | 25 |
| <i>Consumer Reports: Car Buying Guide</i> | 25 |
| eBay..... | 65 |
| FDA Nutrition Fact Panels..... | 113 |
| <i>U.S. News and World Report: America's Best Colleges</i> | 163 |
| Case Study Commentaries..... | 209 |
| Margaret Holmes-Rovner, PhD..... | 209 |
| David E. Kanouse, PhD..... | 225 |
| Stephen Parente, PhD..... | 239 |
| Dale Shaller, MPA..... | 250 |
| Shoshanna Sofaer, DrPH..... | 263 |
| Lessons Learned: Key Variables of Success..... | 275 |
| Advancing Healthcare Decision Aids..... | 293 |
| Getting Tools Used Research Team Biographies..... | 311 |
| Acknowledgements..... | 317 |

GTU research was supported by The Changes in Health Care Financing and Organization (HCFO) initiative, a program of the Robert Wood Johnson Foundation, the California Healthcare Foundation and the Foundation for Informed Medical Decision Making.

Case Study Commentary

Shoshanna Sofaer, DrPH

I. INTRODUCTION

I have been working to develop and evaluate health related decision support tools for the public for about 25 years. My work began with a quasi-experiment to assess a new approach to helping people with Medicare choose among health care coverage options. Four aspects of this first applied study bear examination. First, the dissemination channel was small-group workshops, facilitated by graduate students, and held at dozens of local aging services agencies throughout Los Angeles. We hardly ever think about this approach to providing decision support any more, though perhaps we should. Second, we proactively marketed the workshops, through in person presentations at the same agencies where they were to be offered, as well as others. We got into these settings because we were trusted by the agencies as being both free of conflict of interest and knowledgeable.

Third, our new approach worked. Our experimental group got informational materials based on a method we developed called the Illness Episode Approach. This approach took the complexities of variations in health insurance coverage for people with Medicare and “hid” them by generating a bottom line figure for people – the out-of-pocket costs they would face for a given common illness episode under Medicare alone, or half a dozen other forms of Medicare coverage. Our comparison group got a set of materials available at that time to area seniors that used a more typical chart of variations in premiums, deductibles, co-payments or co-insurance, limitations, etc. Both groups showed significant increases in knowledge; but the experimental group also ended up making changes in coverage that saved them a lot of money.

Finally, no one paid much attention, until quite a bit later. The Health Care Financing Administration (predecessor to the Centers for Medicare & Medicaid Services or CMS) funded the study but did nothing with the results. At that time, the late 1980’s, “beneficiary education” was virtually non-existent in the agency. The approach was used in a modified form in early issues of Consumers’ Checkbook. More recently, a similar approach has been used, both by CMS and Medicare Part D plans to help people estimate their pharmaceutical costs based on current prescriptions, and identify automatically through a computer decision support tool the plan(s) that would result in the lowest out of pocket costs for prescription drugs. These tools have been used extensively by people on Medicare and those who help

them make decisions; unfortunately, sometimes the results you get on the CMS tool are different than those you get using the insurance plan tools.

So. We try something to simplify complex health care decisions for people, it works, and not much happens in terms of adoption of the tool in question. Thus, for me, the question of “getting tools used” has always been compelling. If anything, my subsequent experiences in providing tools to help consumers compare health plans and providers has been even more frustrating. Many designers and researchers have worked hard to figure out what methods work best to make the complexities of healthcare quality more accessible and relevant to our audience. The results of our research are ignored because of bureaucratic and interest-group resistance, especially from those being rated. The reports that have been produced to date are really great in some cases and depressingly awful in others. In virtually all cases, however, they are not getting used by the intended audience – consumers, patients, caregivers, the public. They can still have an impact on providers, primarily because providers don’t like looking bad even if the ratings don’t affect their market share

At the same time, the American public has gotten familiar with using “ratings” of various kinds to help them make decisions. What can we learn from these experiences about what it might take for decision support tools of various kinds to get used (we presume for good) in healthcare? Here’s my preliminary answer based on review of the four incredibly detailed and carefully constructed case studies that this project has generated.

II. INSIGHTS FROM INDIVIDUAL CASE STUDIES

A. *Consumer Reports: Car Buying Guide*

There are numerous lessons to be drawn from the *Consumer Reports* (CR) experience. Furthermore, they have already moved into ratings of healthcare and plan to go further in that area in the future, having created a “Health Ratings Center” and funded it well over the next few years. An important contextual factor for this case, however, is that the organization Consumers Union and its magazine have been around for over 75 years. This provides a foundation that many other potential sponsors of decision support tools would find difficult to match.

Looking at our analytic framework, the most significant factor supporting CR’s success is the fit between the nature of CR’s objectives and the brand it has created. When this is combined with an understanding of its audience and the expertise behind the ratings themselves, i.e., the tool design, we have a pretty powerful combination.

CR has two primary objectives: (1) they want people to buy safer, more reliable automobiles that provide high value at the point of purchase and beyond; (2) they want to sell magazine and Web site subscriptions since that is their major source of revenue. We would need to have more details to discover if these objectives are ever in conflict, but on the surface they do not appear to be. It is not clear to this reviewer whether the “mission” objective and “business model” objective of some health-specific decision tools are typically this well aligned. For example, many quality reporting tools depend on data from the providers themselves; this makes it difficult for report sponsors to call the shots regarding how those data will be scored, displayed, framed, and marketed.

In achieving its objectives, CR’s most important resource is the CR brand and the automatic trust that it brings. The public believes that ratings are totally objective, expert, and independent. An important lesson here is that CR is emphatically NOT taking into consideration the viewpoint of all stakeholders in the automobile market. They are focused entirely on the consumer. It is rare indeed that decision tools generated in healthcare have this level of brand identity, independence, and freedom from pressures from multiple stakeholders with diverse viewpoints. CR also has the resources to support, through their subscriptions and some donations, a complex facility for actually testing cars themselves, as well as an experienced and skilled editorial staff. This balance of technical ability and ability to effectively communicate with the audience is, again, seldom found in healthcare, although that may be changing.

CR as a basic magazine has a large audience which is not representative of the U.S. population as a whole. Their audience tends to be better educated, somewhat more affluent, and until recently has been skewed to the middle-aged rather than the young. CR has had a long time to learn about its audience. It is constantly working to learn more about how the audience is changing, what they perceive they want and need, how they react to changes that are made, etc. They use this information to help shape the content of their products.

However, they also try to influence what their audience wants by focusing on certain issues that they, consistent with their broad “consumer protection” mission, believe have to be taken into consideration. Examples of this include automobile safety and gas mileage. This tends to further shape the audience to those who think those are critical issues, but it does not appear to lose audience share, since they continue to consider other aspects of the car buying decision (such as cost, reliability, handling,

etc.). A key feature of this audience is that it includes a lot of people who are actually considering the purchase of a car, which can happen at any time. The audience knows that the CR ratings are available year-round, in addition to having “new news” with a special issue each year.

With respect to tool design, what is critical is that CR does not depend on data from others for its auto ratings (with the exception of crash test data). They generate their own criteria/measures, they do their own testing and surveying to gather data, and they have generated their own approach to scoring, ranking and rating options. They work hard to justify their approach, but auto manufacturers and dealers cannot pressure them to do it their way. Indeed, when automakers have objected to their ratings, this actually seems to generate media attention, helping promote the ratings.

Other lessons to be learned from the CR auto rating tool’s design reinforce what we have learned in health care: (1) rank ordering ratings is a good thing; (2) people like recommendations from these independent experts as well as information they can use to come to their own judgments; and (3) you can use the decision point as a teachable moment to point out, in editorial content, things people need to pay more attention to and myths they need to get over.

B. eBay

The most significant variable that has led to the phenomenal success of eBay (in spite of recent declines in use), is that it uses Web 2.0, an approach to tool design that depends on active participation of the audience in creating the environment in which people can make decisions. This means the audience is inherently central to the whole enterprise, and eBay’s challenge is to continue to evolve in response to audience preferences and the competition.

Another key variable is that the objectives of eBay as a business enterprise are not just to “educate” or even “persuade” as is often the case in healthcare, but rather to generate purchases. People don’t really use eBay to make decisions; they make decisions, i.e., buy things, as they use eBay. Now this is a “buyer’s” perspective, and another challenge for eBay is to balance the needs of buyers and sellers, since both kinds of parties are needed to make this work. This sounds very utilitarian but in fact, another key factor to the success of eBay is that it works as entertainment. It needs to work as entertainment, especially be more entertaining than the competition (other Web sites, or actually hitting the streets to go shopping). It also needs to provide the

economic benefit of bargains. Buying something brand new in the box with the tags for \$100 when it retailed for \$1000 is truly addictive.

The design of the tool emphasizes efficient and flexible searching for items to compare and purchase. Navigation is key to the effectiveness of this tool. The tool is not perfect by any means. Searching can be incredibly cumbersome, especially, oddly, if you really know what you are looking for and the search terms you use just don't seem to be working. It is, however, good enough, although as competition for the "space" it occupies increases, it will have to become a lot better. Another feature of the design that seems to work is that you can not only shop and buy things on eBay; you can quickly and easily pay for them using PayPal, their affiliate, which also ships many items.

Two highly related variables leading to the success of eBay are that (1) as the pioneer, it has huge brand recognition; and (2) it has created mechanisms to overcome what could have been a complete deal-breaker for the site, i.e., the ability to generate trust in the overall process by helping users identify buyers and sellers who are trustworthy. Again, here, they have not done a perfect job. There are still plenty of knockoffs and poor merchandise on the site, but they have done a "good enough" job to generate billions of purchases. Note that eBay does not say a buyer or seller is trustworthy, other buyers and sellers say so (or do not).

On a personal note, I happen to be an avid eBayer (on the buying side only for now), and it is quite remarkable how one's relationship to the site can become so individualized. I search for specific brands, go first to specific sellers with whom I have done a lot of business, and feel no compunction taking advantage of the ability to ask highly specific questions. I get what I need on the site, and presume that others with *very different needs than mine* can do so as well. Can we build that kind of flexibility, reach, and personalization into health-related tools?

C. FDA Nutrition Fact Labels

This is the only case that focuses on a tool developed by a government entity in response to a legislative mandate. This appears to represent both a significant resource and some degree of constraint. On the one hand, the federal Food and Drug Administration has the power to require manufacturers of specified food products to conduct testing to determine the nutritional content of the items they sell and to place

information about that content on every single food package, in a single prescribed manner/design.

On the other hand, placing this activity in a government agency appears to have constrained the promotion of the label over time, the evaluation of its impact, and the creation of ancillary educational programs that might help a broader audience to actually understand the tool and how to use it.

In part, this appears to be a function of limited financial resources, but it may also be because the nutritional label, I would expect, has to compete for attention within the agency against much higher-visibility activities such as approval of new pharmaceutical products and medical devices.

It appears that the label has become an ubiquitous part of the food shopping experience that is familiar to many and used by some. The extent to which it changes consumer behavior, and whose behavior it changes, is unknown. The effort reminds one of the many mandated “transparency” initiatives that result in the publication of a comparative quality and/or cost report that is never evaluated to see if it makes a difference to anyone, perhaps because no one cares.

D. U.S. News & World Report: America’s Best Colleges

Like the *Consumer Reports* auto ratings, the *U.S. News* College Ratings are embedded in a magazine which has been around for a long time, and focus on an infrequently made but high-price and high-stakes decision. Also like CR, the *U.S. News* ratings are intended to sell magazines and drive visitors to its Web site. However, because it is a traditional outlet, it makes its money by selling advertising, and the readership is key only as it affects the rates they can charge and the number of advertisers; thus its objective is different. All this still means that *U.S. News* needs to be highly attuned to its audience and to the decision context in which its audience operates.

On the other hand, they don’t really care who the audience is, as long as they use the tool and either pay for it or generate other revenue by using it. It has no mission to improve either the quality or value of college and graduate school. Ironically, it has a huge impact on colleges and universities, one which may be truly perverse at times, as when an institution trying to get into the “top 25” of any category turns itself inside out to get there, using strategies that have little to do with the education they provide or even their costs. This appears to be an artifact of the inclusion in the measures of a lot

of “reputational” information from other colleges and universities¹. The degree to which other institutions actually know anything about those they are rating, or would be likely to report it accurately, is debatable.

While its early entry status means that it has a resource a good deal of brand recognition, trust does not seem to be as essential to it. It merrily changes its methodology on a regular basis, something healthcare raters are often terrified to do. Oddly, these changes lead, just as do the controversies over CR’s auto ratings, to media coverage and thus promotion of the ratings. U.S. News & World Report interacts with those it rates, and has responded to their complaints. This may be because the institutions are essential to the collection of much of its data.

Note that the magazine does not pretend to be the primary source of information for people making a decision. Rather, the tool is designed to help people narrow their choices, check out choices they have already narrowed down to or even made, and find in one place a lot of information some of which might be of interest to them. Finally, U.S. News puts enormous energy into promoting the college ratings and expanding the methods by which they are disseminated. They view this as a cost of doing business and essential to their strategy (selling magazines and getting people to the Web site) rather than a luxury item.

III. CROSS-CASE ANALYSIS

Analyzing across the cases leads to the identification of several key variables, some of which are not a part of our conceptual framework. One of these appears to be related, as well, to our criteria for selecting tools to examine. That variable is the national scope of each of these efforts. Both *Consumer Reports* and *U.S. News & World Report* ratings present information on products and institutions, respectively, across the country. Note that the number of such entities appears not to be a barrier to presenting data on almost everyone.

The FDA label exists across the country and covers all American, and some foreign, products and manufacturers within the legislative mandate. eBay is a national enterprise – it displays items from sellers across the country (indeed around the world) to buyers similarly dispersed.

¹ A disclosure here: I am a faculty member in a College and a School that is rated by U.S. News. My School, in particular, is currently rated *solely* on the basis of reputational data, since there isn’t any other standard data.

The market for automobiles is a national market. While many colleges appeal primarily to a local or regional pool of potential students, others reach out to students from across the country. The market for many of the products on eBay is also national or international in scope. Some products may have a narrower appeal (shearling-lined gloves might not be a hot item in South Florida) but most have a broad appeal. Food is ubiquitous; while there are regional manufacturing operations, and while comparisons and purchases are mostly made at the local level, you can find Campbell's soup almost anywhere.

In contrast, healthcare is local. While someone with a rare or very serious disease may consider going across the country for medical care, people typically stick close to home. While there are some healthcare decision tools that are national in scope (the CMS "Compare" sites come to mind immediately), most comparative quality reports are quite localized. We will discuss the implications of this in our final section.

Second, all of these tools have a strong brand identity, whether they have been around for a long time or just a decade or two. Except for the FDA, they all have competition, but they have a substantial share of their market. Brand identity does not appear magically or overnight. It requires considerable sophistication in both designing a tool with an audience, decision context and function clearly in mind, and considering it completely normal to spend considerable resources on promotion, dissemination, testing, and evaluation. The exception, the FDA, is a government agency, and government agencies do not seem to be comfortable with the idea of marketing themselves or being terribly transparent about how well they are actually doing (in spite of numerous recent efforts to require federal agencies to self-evaluate).

Another shared characteristic of the three non-governmental sites is that they do not seem to worry a lot about exactly who their audience is, as long as they have one that is large enough and has enough financial resources to help them meet their objectives. Thus, in spite of its avowedly social mission, CR does not seem to worry a lot about not reaching "vulnerable populations;" U.S. News & World Report similarly is not working hard to reach young people just back from Iraq and Afghanistan and heading to college, or people coming back to school in their 30's or 40's. Even the FDA seems not to care that a high proportion of the population simply cannot do a lot of the math required by its label, even in terms of figuring out how many "servings" can be found in a package of eight Oreo cookies.

Reviewing these cases made me rethink the issue of trust. We take it as given, and truly believe, that a decision support tool in healthcare will not be used unless it is trusted. In my own research, I have learned that trust seems to depend on two factors: belief that the people presenting the ratings have the necessary technical expertise to get them right and belief that

they are independent, i.e., do not have a conflict of interest. It appears that for all these tools, users trust the sponsor enough to use the tool; but that the depth of trust and its basis varies.

Consumer Reports' success is largely build on trust. They are independent, they are expert, and they are clearly in the consumers' corner and no one else's. That is their brand, that is their mission. I venture to say that their audience would not consider anyone else more trustworthy when it comes to automobiles. The fact that they are trustworthy is actually reinforced when they get attacked by manufacturers.

The FDA labels seem to take trust for granted. Frankly, if I thought about it at all, I presumed that the FDA came up with the information on the labels for each product, but of course, that's not true. The manufacturers do. Yet we just trust the label because it is an FDA label (because it is a government agency? because it is the FDA in particular and we trust them to approve pharmaceuticals?). Is the trust justified? Does the FDA check on what the manufacturers are saying often enough and rigorously enough? Does anyone care about that as they stand in the supermarket aisles comparing the calorie count of two different cans of chicken broth?

eBay works very hard to create a safe environment for buyers and sellers. It would not succeed if people were too wary to sell things and buy them. The "feedback" mechanism is a central part of this and it has been changing lately, with some disgruntlement from sellers. Yet even eBay refuses to take total responsibility for the trustworthiness of its participants. If someone sells a Gucci knock-off and someone else buys it, eBay does nothing. If the listing does not say "authentic" and the buyer doesn't ask about authenticity and get a positive answer, then they are likely not to get much help even from the "dispute resolution center." After all, some people *want* to buy Gucci knock-offs, as long as they are well-priced, reasonably well-made Gucci knock-offs.

Similarly, U.S. News & World Report would like to be trusted, but they probably don't need total trust from readers, who usually expect to use other sources of information in making the decision. The notion that some but not complete trust may be adequate for getting tools used is, frankly, surprising.

IV. IMPLICATIONS FOR HEALTHCARE

Should healthcare decision tools be national in scope?

Currently, healthcare decision tools are developed and disseminated primarily for local audiences. Typical sponsors are state agencies, purchaser coalitions, multi-stakeholder groups, and the like. This decentralized structure has disadvantages that are highlighted by the cases. These include:

- No sponsor seems to have enough resources to do adequate audience research; design, and formatively test an evidence-based report; promote it effectively (which requires not a one-time push but ongoing efforts); evaluate it and adapt it in response to evaluations and changing audience needs; and create a strong brand identity.
- At the same time, each one of these sponsors is struggling to get it right, and often re-inventing wheels, some of which turn out to be square.

Are there advantages to decentralization, or perhaps better put, are there serious disadvantages to centralization? The major disadvantage to centralization is that we have to put all, or almost all, our eggs in one basket. Designing appropriate guidance and governance for such a single central sponsor would be difficult. The experience of the FDA in terms of their inability to focus resources on the food label as it competes with many other agency priorities is of concern. Of even greater concern is the experience of watching the one centralized reporting effort we have, that of CMS, be constrained by internal and external politics, including the conflicting goals of different, and powerful, stakeholders. This leads to a second question.

Should healthcare decision tools be produced only by entities who have a clear commitment to the consumer/patient/public and to no other stakeholder?

Three of the four sponsors in these cases appear to be freer of pressures from those rated than are current sponsors of healthcare decision tools. There is variation on this across the cases. Consumer Reports is very independent; the FDA appears to have operated with considerable independence; U.S. News & World Report does adjust its scoring and ranking methods in response to pressure from college and universities, but the method is still controlled by them; and I suspect that they would not give in to pressures to change the ratings in ways that would lead to fewer sales. eBay is by its very nature created by participants, but even there the “rules of the game” such as they are, are set by eBay, although they are open to a lot of feedback from members. Outsiders, such as luxury goods manufacturers LVMH, have had little luck in getting them to act as police in their interest.

This freedom allows the sponsors to focus on their audience in all aspects of their operation. This in turn, it can be argued, makes it far more likely that they will reach and keep their audience if (and this is a genuinely significant if) they are able to come up with a meaningful product/service. In healthcare, pressure from other stakeholders makes it difficult for sponsors to even identify their audience clearly, let alone meet their needs without worrying about what others will think. A whole series of evidence-based recommendations are regularly resisted and ignored because of pressure, typically from the providers being rated. Examples include rank ordering providers by performance from best to worst in displays, rather than using alphabetical order, and pointing out significant gaps in performance to encourage people to look at the report. .

A caveat is needed here. The sponsor needs to have expertise that is far beyond that of the typical audience member. This is, after all, about knowledge transfer and there has to be a knowledge base. CR's automobile ratings are based on the existence of a complex and sophisticated facility and staff to actually test automobiles and survey automobile buyers about their experiences. The FDA label required extensive technical expertise. One might wish there were more technical expertise available at U.S. News & World Report, but over the years they have learned (and the field has also become far more focused on performance measurement and transparency). Thus, I am not suggesting that Angie's List or Zagat is the answer for healthcare ratings, for example, although it will be interesting to see what happens as they expand into physician ratings.

We may want to consider, however, whether an independent entity, whose board is made up exclusively of people committed to the interests of healthcare consumers and patients, and whose staff includes a wide range of experts in measurement, report design, marketing, and evaluation, is not what we really need in healthcare.

Perhaps such an entity can also revisit the issue of what the actual *product* is that consumers and patients, or more likely specific subgroups of consumers and patients, would really find useful. Many believe that our current ratings do not, at least yet, speak to the currently felt needs of consumers and patients, who tend to believe that their providers are wonderful even if other people's providers are not, and that although health plans are distrusted, they are all the same so which one you choose makes no difference. In contrast, people do *not* believe all automobiles are alike, or that all colleges are alike, that all cookies have the same number of calories, or even that different sellers on eBay will give you about the same price and service when you buy a pair of "Sex in the City" shoes.

The other question that remains is - what's the business model here? Is this a public utility that gets funded by a small percent of healthcare revenues? Should it have core funding from government but be required to generate revenues from users as well, perhaps in the manner of public TV and radio? This leads to our third concern - the level of resources available for these efforts.

Can we get serious about getting tools used?

To date, our approach to the work of designing and distributing decision supports in healthcare has been a highly marginalized activity which is not only woefully underfunded, but which no one expects would ever be other than woefully underfunded. We have to ask the question - why bother to waste our time and energy on an unreachable goal?

In each of the cases, sponsors expected that they would have to use resources to learn about their audience, select an appropriate audience segment, design and test a sophisticated and useful tool with their needs in mind, promote the tool intensively with a clear goal of getting it used, evaluate their work regularly, and make changes in response to both evaluations and changes in the context and the audience. It is likely that getting tools used in a meaningful way is going to require us to recognize that this will not be done on the cheap.