12-2014

Effects of Video Enhancement in a Stated-Choice Experiment on Medical Decision Making

Susanne Hoffmann
Ludwig Maximilians Universitat, Munchen

Joachim Winter
Ludwig Maximilians Universitat, Munchen

Francis G. Caro
University of Massachusetts Boston, frank.caro@umb.edu

Alison Gottlieb
University of Massachusetts Boston, alison.gottlieb@umb.edu

Effects of Video Enhancement in a Stated-Choice Experiment on Medical Decision Making

Susanne Hoffmann[1]

Joachim Winter[1]

Francis G. Caro[2]

Alison S. Gottlieb[2]

[1] Department of Economics, Ludwig-Maximilian University, Munich, Germany

[2]Gerontology Institute, University of Massachusetts-Boston, Boston, MA 02125, USA

Abstract

Background. The internet can be useful in administering stated-choice experiments to understand medical decision making and refine the content of patient decision aids. In internet-based stated-choice experiments, video and audio files can be used to provide information to respondents. Quality of data may or may not be affected.

Objectives. In a methodological experiment concerned with administration of a stated-choice experiment on the internet concerned with knee-replacement surgery, we compared the data quality obtained with video-enhanced and conventional text formats.

Methods. Members of the RAND Corporation's American Life Panel and 50 years of age or older (n=1616) were randomly assigned to one of two survey modes: video (80%) or text (20%). Quality of data was measured by both the frequency of respondent speeding and plausibility of respondent choices.

Results. Typical respondents completed the research task much more rapidly in the text condition than the video condition (10.6 versus 16.4 minutes). The choices were 5 to 10% more likely to be plausible for those in the video-enhanced condition than for those in the text version. The mode of administration did not affect respondents in the extent of their reported interest in the survey or in the degree to which they reported identifying with patients in vignettes.

Conclusions. The advantages of the video enhancement in engaging respondents appeared to be offset by the greater time demands on respondents. Video-enhancements in stated-choice experiments may be most useful when used selectively to deliver information that cannot be delivered adequately by text.

Acknowledgements

The research was supported in part by investigator-initiated grant number Investigator initiated grant: 0212-1 from the Informed Medical Decisions Foundation. Maria E. Suarez-Almazor, MD, PhD served as medical consultant for the project. Arie Kapteyn, PhD, Tania Gutsche, and other staff members of the RAND Corporation's American Life Panel arranged for production of the video files, programmed and administered the experiment, and prepared the data files.

Effects of Video Enhancement in a Stated-Choice Experiment on Medical Decision Making

Introduction

To engage respondents fully in a stated-choice experiment concerned with medical decision making and administered on the internet, we have made use of video and audio formats in delivering information. Stated-choice experiments can provide an efficient means of approximating actual experiments [1]. The quality of stated-choice experiments depends on engaging respondents in the experimental task. Our work on these methods reflects recognition that many people are now accustomed to receiving extensive information through motion pictures and the spoken word rather than through written text. Well established experiences with theater, motion pictures, television, and the internet make it clear that video and sound also have a capacity to deliver information and emotional content in ways not readily captured by written language. We have demonstrated that it is technically feasible for researchers to administer complex stated-choice experiments on the Internet with extensive information provided through video clips [2].

In this paper, we examine whether there are advantages to delivering information in a stated-choice experiment through pictures and the spoken word rather than through text affects data quality. We report the findings of a methodological experiment that compared the data quality obtained with a video-enhanced format with the one obtained with a conventional text format in an online vignette experiment. The vignette experiment was substantively concerned with knee-replacement surgery as treatment for osteoarthritis of the knee. The main substantive findings of the study have been reported elsewhere [3].

Background

Stated-choice experiments are used in a variety of fields including health care, environmental studies, market research, and ethics to understand the basis of complex decision making [4, 1, 5]. Terms used to identify stated-choice methods include “conjoint analysis,” “vignette experiments,” and “contingent

valuation" [6]. In health care, stated choice methods can be used effectively to study patient decision making. These methods can also be used to provide a basis for refining the patient decision aids that are now widely used. According to International Patient Decision Aid Standards (IPDAS) [7] over 500 decision aids are available or being developed world wide. The internet is now being used to make decision aids available to patients and some studies are appearing concerned with the effectiveness of decision aids administered on line [8]. Consequently, there is good reason to examine the effectiveness of the internet in delivery of health information to patients.

In stated-choice experiments, respondents are asked to make judgments about hypothetical situations described to them through vignettes, that is, very short stories. Systematic manipulation of vignette content allows investigators to determine how specific elements within vignettes affect judgments. Respondents must be able to grasp the vignette content quickly and identify with the situation described to them. A well-established approach to engage respondents is the use of a "vignette person" whose situation respondents are asked to consider [4] .

In a self-administered stated-choice experiment, text is the most basic option for delivering information to respondents. In stated-choice experiments conducted through face-to-face interviews, interviewers can combine oral presentation with text and pictures. In telephone interviews, respondents usually have access to questionnaire content only through the language read to them by interviewers.

In administering stated-choice experiments, investigators may experience a number of advantages by providing information through pictures and spoken words rather than by text alone. Visual images and human speech may also convey emotions more efficiently and sometimes more powerfully than written language. Furthermore, images and speech may be useful in engaging respondents' interest and are able to reach those who do not read or simply prefer to receive information through pictures and spoken language. On a more methodological note, pictures can also include information whose effects an investigator may want to measure unobtrusively. Pictures, for

example, have been used to study racial prejudice. In these studies, information about race and ethnicity has been introduced only through the appearance and perhaps speech patterns of the actors [9]. In a contingent evaluation study concerned with an environmental issue, Mitchell and Carson [6] used photographs in face-to-face interviews to help respondents grasp the potential visual impact of construction of a power plant in an environmentally sensitive area.

Although not yet widely recognized, the internet makes it possible to include pictures and sound in administering stated-choice experiments. The use of the internet to administer traditional text-based questionnaires in a paperless format is well established [10]. In a recent review article, Couper [11] provides an overview of the manner in which the internet has expanded options for investigators in administering surveys. Among the options supported by the internet are the uses of both visual images and recorded sound in administering questionnaires.

Use of video and audio technology in administering stated-choice experiments also introduces challenges and possible disadvantages for investigators. One barrier is the expense associated with producing videos. Preparation of scripts, identification of locations for shooting videos, preparation of the settings, recruitment of actors, coaching of actors, and recording the performances are all likely to be sources of costs. The preparation of high quality audio recordings is a distinct and potentially costly challenge. Formatting video and audio files so that they can be seen on computers with various operating systems and software for viewing video files is another technical challenge.

Investigators have opportunities to convey information through the performance of actors, their clothing, and the settings. At the same time investigators are challenged to avoid inadvertent inclusion of pictures and sounds that detract from the research agenda.

In this paper we report on the findings of a methodological experiment concerned with an elaborate use of videos in internet-based research. (The methodological experiment was conducted within a stated-choice study that itself was an experiment.) We generated subsamples randomly and

assigned them to either a video version of our survey or a text-only version to study the effects of the video survey on various measures of survey quality. Our study is one of the first to examine methodological aspects of the use of videos in internet-based surveys. A precedent to our study is the paper by Fuchs and Funke [12] in which a text-based survey was compared to a survey in which a video was used to permit respondents to see an interviewer who read questions to them.

Methods.

Study sample. Study participants were members of the RAND American Life Panel (ALP) who were ages 50 and older at the time of the survey and had not participated in our pilot study [3]. Members of the ALP are drawn from the general population and are surveyed periodically in an internet format. Respondents without internet are provided internet access by RAND. They receive modest financial compensation for participating in particular studies.

The ALP invited participation from all 2296 of its members who were 50 years of age and older and had not taken part in a pilot version of the study. A total of 1675 interviews were started; 1622 were completed. Among panel members, the response rate was 70.6%. Background characteristics were incomplete for six respondents, who were subsequently dropped from the analysis. This resulted in a total respondent sample size of 1616.

We had access to standard demographic data about respondents that had been obtained when they joined the panel.

Study design. Respondents were asked to make recommendations regarding knee-replacement surgery for three hypothetical patients (vignette persons) whose circumstances were described to them along several dimensions. These included the patient's pain level, his or her opportunity costs associated with having surgery, the recommendation of the patient's surgeon, a second opinion, and an outcome forecast. In some cases, respondents also had access to a testimonial from another patient. We used a

fractional factorial research design. Randomization was used to determine which dimension levels would be included in specific vignettes for specific respondents.

To provide the basis for the examination of the effects of the videos on survey quality, respondents were randomly assigned to one of two survey modes: video (80% of respondents) or text (20% of respondents). In the video mode, most of the information was conveyed through videos in which professional actors performed as patients and physicians. In the text mode, exactly the same information was provided, but in written form. To enable all study participants to complete the survey, we provided respondents in the video mode with opt-out options if they experienced technical problems in watching the videos or listening to the audio tracks. Switching from the video to the text mode was entirely at respondents' discretion. Consequently, there may have been other reasons for respondents switching from the video to the text version. In the analysis section, we label this group the "text-fallback" group.

Survey administration. At the start of survey administration, we provided an audio and video setup check. We then asked respondents to judge a vignette on an unrelated topic to gain experience with the research process. We administered the knee-surgery decision experiment as follows: First, a mini lecture by a physician on knee osteoarthritis was provided. Respondents were then asked to review the situation of a vignette person who was considering knee-replacement surgery to treat knee osteoarthritis. So that respondents would receive the vignette person's presentation seamlessly, all possible combinations of each vignette person's characteristics were video recorded. Respondents then received information about a surgeon's recommendation, a second opinion provided by another physician, and a patient-specific outcome forecast. For the third vignette, most respondents also had an opportunity to consider one or two testimonials from patients who had received knee-replacement surgery. The outcome forecasts were provided to all respondents either through text or graphs. Before making their recommendations, respondents had the opportunity to review a written summary of the

information provided to them on all of the dimensions except for the testimonials. The summary was provided to minimize possible recency effects [13].

After considering the vignette, respondents were asked to recommend whether or not the vignette person should have knee-replacement surgery using the following question:

Do you recommend that **Name** have full knee-replacement surgery now?

____ Yes

____ No

The respondent's recommendation was always elicited in a text format.

After making a recommendation on the first vignette, respondents were asked to consider two more vignettes.

Analysis

Description of the study subsamples. We report findings for three groups of respondents: "assigned text," which included all respondents randomly assigned to the text mode, "text fallback," which included all respondents originally assigned to the video mode but used the provided opt-out options, and "video," which included the remainder of the respondents originally assigned to the video mode. The "text fallback" group was composed as follows: 229 respondents who failed the video setup and 129 respondents who requested the text version during the vignette person's video and the surgeon's recommendation for at least two vignettes. (With this definition, we ensured that respondents were assigned to the text fallback group only when they had requested the text versions for a large part of the survey. Most of these respondents also requested text versions for statements by all three vignette persons and for the recommendations of the surgeons.)

In total, 358 respondents were reassigned to the "text fallback" group. Table 1 describes the respondents according to the survey format through which they participated in the study.

[Table 1]

While there were no striking differences in demographic characteristics between the "assigned text" and "video" subsamples, the "text fallback" subsample had age and employment profiles that were different from the video subsample. The “text fallback” subsample was older and consequently also had more retired respondents. A larger fraction of these respondents in the “text fallback” subsample were also in the low-income category (which might also be related to age and retirement). We also found differences in ethnicity and marital status. For the older respondents in this subsample, viewing the video files was more likely to be a challenge than for younger respondents. However, a substantial number of respondents in the “text-fallback” group were younger than age 60. Clearly, difficulty with the videos was not simply explained by respondent age. Since the "text fallback" subsample is linked to a set of respondent demographic characteristics, we distinguished between "assigned text," "text fallback," and "video" subsamples throughout our entire analysis.

Survey quality measures. We had two bases for comparing the implications of the survey mode: measures of respondents' survey behavior and assessments of the respondents' survey experience. The measures for respondents' survey behavior are the following: Item nonresponse, survey duration, time needed for decisions, speeding behavior, and a measure for plausible decisions. Item non-response has long been established as an important indicator of data quality [14]. Item non-response results in a loss of information, affecting the efficiency of estimates and the power of statistical tests. Furthermore, if item-nonresponse is associated with respondents' characteristics, it may lead to biased results [15]. Lengthy survey duration may increase respondent fatigue and burden. Thus, the time it takes respondents to complete the survey is also likely to affect survey quality. The time respondents need to make a decision can be used as a measure for respondent burden [16]. An increased respondent burden can affect the study quality adversely. The extent of speeding behavior, that is, the extent to which respondents proceed through the survey extremely quickly, can also be an indicator of data quality. This

may be especially crucial for stated-choice studies, where it is essential that respondents fully grasp the vignette content. If respondents spend extremely little time on reviewing the given information, this raises the suspicion that they did not pay full attention to the provided content and that the judgment on the vignette they provide may be flawed by this behavior. A measure for the plausibility of decisions is included in this study as a direct measure of the quality of the most important answers from respondents, the vignette recommendations. If the mode of delivering the information affects the quality of these responses, this might undermine the validity of the results obtained with the stated-choice experiment.

The assessments of respondents' survey experiences were elicited at the end of the survey. We asked respondents to rate the degree of interest in the survey and the degree of identification with the vignette persons. The answers to these items serve as indicators of respondent involvement in the survey.

Results

Item nonresponse. Item-nonresponse was a minor issue. For all surgery recommendations, we missed a response in only three instances. For the 16 questions about respondent experience with knee osteoarthritis and knee-replacement surgery, 92% of the respondents answered all of the questions. Another nearly 8% answered all but one question. None declined to answer more than two questions. Although nonresponse was rare, the rate of nonresponse was significantly higher in the "text fallback" group, indicating a lower data quality in this subsample.

Survey duration. Respondents varied greatly in the time they spent reviewing vignette content and making decisions. Table 2 presents the mean time (measured in seconds) spent on the entire survey for the full sample and the three subsamples. Because distribution of the time spent on completing the entire survey was heavily skewed with a long right tail, we used quartiles to measure central tendencies and variation. Consideration only of those who fell between the 25th and 75th quartiles eliminated the

distortions caused by respondents who were either very rapid or very slow in completing the protocol. For the full sample, the median time spent on the survey is 905 seconds or approximately 15 minutes. [Table 2]

For the subsamples, we found the following: The assigned text subsample completed the research task the most rapidly; the median time was 10.6 minutes. The median time of completion for the "video" subsample was approximately 6 minutes greater (16.4 minutes) than the median for the "assigned text" subsample ($p<0.01$). (The p-values in this subsection are based on non-parametric k-sample tests for equality of medians.) The median for the "text fallback" subsample was a little lower than the median for the "video" subsample ($P<0.01$) and higher than the median for the "assigned text" subsample ($P< 0.01$). Therefore, members of the "video" subsample may have experienced a higher respondent burden and their responses may have been subject to increased fatigue because of the longer survey duration.

[Table 3]

Examination of time spent on various vignette components is helpful in understanding the differences in time spent among the three subsamples. The distribution of the time spent on the single components is very similar for the "assigned text" and "text fallback" subsamples, and the median time spent in these subsamples is considerably lower than time spent for the "video" subsample ($P<0.01$). The only dimension where the median time spent in all three subsamples was fairly close was the outcome forecast dimension. This similarity can be explained by the fact that information for the outcome forecast was displayed in a text format for all subsamples and was only supported by a brief audio narration for the "video" subsample. However, the "video" subsample's median is still significantly higher than the median of the other subsamples for this dimension ($P<0.01$). The variation of the time needed for various components reflects differences in length of the scripts. For the respondents in the "assigned text" and "text fallback" subsamples, additional variation may be attributable to differences

among respondents in reading abilities and reading strategies. For the “video” subsample, some variation is attributable to differences in internet connections and computer hardware, which led to differences in the time it took to load the videos.

Time needed for decisions. Given that the preceding analysis revealed longer survey duration for the "video" subsample, the natural follow-up question is whether this longer survey duration resulted in an increased respondent burden. Bassili and Scott [16] argued persuasively that response times in surveys can be taken as an indicator of respondents' cognitive burdens. Table 4 provides an overview of the total seconds respondents spent on the decision screen for each vignette task for the full sample as well as the three subsamples.

[Table 4]

The distributions of the time taken were similar across the three subsamples, and there were no significant differences in the median time spent on the decision screen. Therefore, even though for the "video" subsample respondents needed significantly more time to complete the survey, we did not find an indication of an increased respondent burden for this subsample.

Speeding behavior. If respondents spend very little time on the information provided; that is, they speed through the survey and may not fully consider the provided information. For stated-choice experiments, it is especially important that respondents fully grasp the information provided so that their choices reflect valid judgments of the situations described in vignettes. Therefore, speeding by respondents can negatively affect data quality. In this section, we examine the extent of speeding. We established criteria for the minimum amount of time a respondent needed to comprehend the information provided. Accordingly, we classified those who fell below the minimum as “speeding.” For the reviews of the personal situations of the patients in the vignettes, an observation was classified as “speeding” if the time spent on reviewing this information was 10 seconds or less. We used this criterion for both video and text versions. All of the videos in which the patients described their situations had

running times of more than 10 seconds. The shortest text description of the personal situation of a patient contained 110 words. Assuming an average reading speed of 250 words per minute [17], respondents could not have completely read through the provided text for the descriptions of the personal situations in less than 10 seconds. For the reviews of the surgeon's recommendation, an observation was classified as speeding if the respondent spent 5 seconds or less reviewing the information. (The threshold reflects the amount of information presented in the surgeon's recommendation.) With the same argument, the thresholds for speeding behavior were also set to 5 seconds or less for reviewing the second opinion, the outcome forecast, and each of the patient testimonials.

Table 5 presents the percentage of observations classified as speeding for the full sample and the three subsamples for all dimensions. In no case was speeding a serious concern. In every category, less than 10% of the observations were classified as speeding. In some categories, less than 1% of observations were classified as speeding. Across the subsamples, the lowest incidence of this behavior was found for the "video" subsample. In most categories, the difference from the other subsamples was significant at least at the 5% level. Thus, while there was little evidence of speeding overall, speeding occurred most rarely in the "video" subsample.

[Table 5]

Plausibility of decisions. Even though there were no correct or incorrect choices, we distinguished between the situations in which a positive recommendation was more plausible and less plausible. On this basis we rated the quality of respondent decisions. Due to the randomized assignment of dimension levels to the single vignette persons, we were not able to classify the plausibility of the surgery recommendation for every vignette. For a substantial proportion of vignettes, we classified the vignettes as more plausible or less plausible surgery candidates, depending on the extent to which the dimension levels included in the vignettes were favorable for a surgery recommendation. In deciding

which dimensions should be used in the generation of this measure, we relied on the main findings of our substantive paper and included the most important decision drivers: pain level, opportunity costs, and physician's recommendations. We then classified a respondent's decision as more plausible if it corresponded with our judgment.

Thus, the classification of plausibility of decisions was based on the following logic: If a vignette person was characterized by high levels of pain, low levels of opportunity costs, and a positive physician's recommendation, this vignette person was classified as a more plausible surgery candidate. Consequently, if this vignette person was recommended for surgery by the respondent, we rated the decision as "more plausible." If not, we rated the decision as "less plausible." Alternatively, if a vignette person was characterized by a low level of pain, a high level of opportunity cost, and a neutral physician's recommendation, the vignette person was classified as a less plausible surgery candidate. Thus, for these vignette persons, respondents made a more plausible decision when they did not recommend surgery and a less plausible decision when they did recommend surgery. Vignettes that could not be classified on their plausibility for surgery were dropped from the subsequent analysis. Applying the above classification to the vignette sample, we obtained 312 vignette observations that were more plausible surgery candidates. Of these observations, 241 were recommended for surgery and are thus vignette observations where more plausible decisions were made. With the same classification, we obtained 895 vignette observations that were classified as less plausible surgery candidates, of which 604 were not recommended for surgery. Thus, in total, we established 845 more plausible decisions and 362 less plausible decisions. Therefore, the clear majority of decisions (70.0%) made for these vignette observations were more plausible. However, the percentage of more plausible decisions varied across our three subsamples. The rate of more plausible decisions was the highest in the "video" subsample (73.1%), followed by the assigned text subsample (68.8%). The lowest percentage was observed within the "text fallback" subsample (63.3%). The difference between the percentage of more plausible

decisions between the “video” and “text fallback” subsamples is significant at the 1% level. This difference remains significant when we control for the personal characteristics of the respondents. Thus, if we take this measure of more plausible decisions as a measure for the overall quality of decisions in our survey, the “text fallback” subsample exhibited a lower quality of decisions than the “video” subsample.

Respondents' self-assessments of the research experience. We used two assessments of the survey provided by respondents to examine whether our use of videos improved respondents' survey experiences. These assessments are of respondent interest in the survey and the degree of identification with the vignette persons. For both measures, a more favorable value in the “video” subsample would indicate that the videos were successful in engaging respondent interest. The assessment of the interest in the survey was measured by the American Life Panel using their following standard question: “Could you tell us how interesting or uninteresting you found the questions in this interview?” Respondents were asked to choose among five options ranging from “very interesting” to “very uninteresting.” The degree of identification with the patients in the vignettes was elicited by us in a follow-up question using the following wording: “Overall, during the course of this interview, to what extent were you able to identify with the people described in the survey?” Respondents were asked to choose among four categories, ranging from “very” to “not at all.”

A majority of respondents rated the survey as “very interesting;” another 40% rated the survey as “interesting.” Only 2% rated the survey as “uninteresting” or “very uninteresting.” The mode of survey administration was not associated with respondent rating of their interest in the survey. Respondents generally reported that they were able to identify with the patients described in the vignettes; 75% of respondents stated that they could at least “somewhat identify” with the patients. Again, the mode of survey administration was not associated with the extent to which respondents reported that they could identify with patients described in vignettes. In sum, the survey generated

high respondent interest and a high degree of identification with the patients in the vignette persons for both the video and text versions. We did not find any evidence of differences between the video and text versions in the extent of respondent interest in the survey or in the degree to which they reported identifying with the patients in the vignettes.

Discussion

Our aim in conducting the methodological sub-experiment reported here was to determine whether the extra effort involved in using video enhancements in an internet-based stated-choice experiment resulted in superior quality data. On the various measures reported here, we found very similar results with the text and video versions. Response rates on the choice tasks were very high on both the text and video versions. The vast majority of respondents in both the text and video versions provided a recommendation for every vignette and answered all other survey questions. Respondents in the video group took considerably longer to finish the survey; however, we do not find evidence that the greater time required for the video version was more burdensome for typical respondents measured by the time needed for decision making. Speeding, that is, implausibly rapid movement through the survey, was not a serious issue in either the text or video groups; however, we found less of this behavior in the "video" subsample. On our measure of the quality of decisions, we rated approximately 70% of the decisions on a carefully selected subset of vignettes as "more plausible." The percentage of more plausible decisions was slightly greater for those in the "video" group than for those in the text fallback group. Since the "text fallback" group is clearly a selected group, this difference cannot be interpreted as a sign of a better decision quality in the video group. There was no significant difference between the video and assigned text groups; thus, we did not find evidence of a better quality of decision in the video group. In addition, the assessments of respondents of their interest in the survey content and the degree of identification with our hypothetical persons was high for both survey versions, but the mode of survey administration did not affect their ratings.

In sum, we were not able to establish striking improvements in survey quality by delivering the information of our discrete choice experiment with videos. One possible explanation for this "non-finding" is that those in the “video” subsample spent much more time completing the survey than those in the other subsamples. While we found no evidence for an increased response burden in the “video” subsample, a longer survey duration has long been established as reducing respondent motivation [18] and could thus have affected the self-assessments and other quality measures so that a beneficial effect of the video mode could not be detected.

The finding of equally high quality data for the text and video versions is encouraging for investigators using text-only methods to deliver information to respondents. The results suggest investigators who limit themselves to text in providing information in an internet-based stated-choice experiment can achieve high quality data without the extra effort involved in using videos.

Our experience has practical implications for other investigators who want to use video-enhancements in surveys. One consideration is the time required for delivery of information by reading text aloud. Respondents who are fast readers can probably complete a survey task substantially more rapidly when they read text themselves than when they listen to information. Investigators have reason to weigh the benefits of delivery of information out loud against the greater time demands implicit in reading information out loud to fast readers. If respondents experience the content as interesting and the delivery engaging, they may not react negatively to the demand on their time. In planning an internet-based study that uses audio and video files, investigators should attend to the total amount of time that will be required of respondents to watch and listen. To minimize time demands on respondents, investigators may choose to use video and audio only on portions of an internet survey where there is a distinct rationale to do so.

When they incorporate video enhancements in internet surveys, investigators should also anticipate that some respondents will not be able to use the video version. Some of the reasons are

beyond the control of investigators. Some respondents, for example, may not have speakers attached to their computers; others may not have the necessary audio or video software; others may have difficulty in adjusting sound levels; some may have slow internet services that will result in frustrating delays in the loading of video clips. Inclusion of a "text-fallback" option is a useful strategy for retaining these respondents. In our case 22% of those assigned to the video enhancement group used the "text-fallback" option. The "text-fallback" option was probably also useful for retaining some respondents who were impatient with the pace of the videos. Since those in the "text-fallback" group differed from other respondents on several demographic variables, the inclusion of the option helped to protect against selection bias. Our experience suggests that other investigators who use video enhancements should also include a text option for those who cannot or will not use a video-enhanced protocol to protect against selection bias.

Furthermore, there are other reasons for using video files unrelated to data quality considerations. Videos may be useful in special instances in which information or emotional content can be conveyed more effectively with visual images and sound than it can by text. Video can also be useful in reaching populations with limited reading ability. Further, videos can be useful in studies concerned with effects of various forms of stereotyping. Videos in stated-choice experiments make it possible to provide information effectively on stereotyping variables such as age, race, and gender, but less obtrusively than through text, and thus to study the effects of these variables on decision making very discretely. In these cases, the additional advantages of using videos may outweigh the associated costs and efforts.

Limitations. We examined the effects of a very extensive use of videos in survey administration by conducting a complex vignette experiment in which almost all information was provided with videos. It may be that a more selective use of videos would have resulted in improved data quality. With more selective use, videos would have had less effect on total survey duration and less adverse effect on

respondent interest. In sum, with carefully selected use, videos may have greater positive effects on data quality.

Because our respondents were all accustomed to completing text-based internet surveys, they may have been engaged with relative ease in a text-based task. A sample more representative of the population with respect to survey experience may have responded with relatively greater positive response to video presentations.

Conclusion. Ours is the first study comparing the relative effectiveness of text and video in internet-based stated-choice experiments concerned with health choices. The potential for generalizing the findings to other situations in which investigators might make some use of video formatting in administering stated-choice experiments on the internet is unknown. More methodological research is needed examining relative effectiveness of text and video formats for stated choice studies on subjects with different age and educational characteristics. In addition, other investigators might design experimental "treatments" that take fuller advantage of the possibilities offered by the internet to deliver information through audio and video formats. These investigators might find that formatting in delivery of information makes differences that we did not observe.

References

1. Louviere, J., Hensher, D., & Swait, J. *Stated choice methods: analysis and application.* New York: Cambridge University Press. 2000. ISBN 9780521788304
2. Caro, F., Ho, T., McFadden, D., Gottlieb, A., Yee, C., & Winter, J. Using the Internet to Administer More Realistic Vignette Experiments. Soc Sci Comput Rev. 30 (2), 184-201. 2012 DOI: 10.1177/0894439310391376.

3. Hoffmann, S., Caro, F., Gottlieb, A., Kesternich, I., Winter, J. Contributions of second opinions, outcome forecasts, and testimonials to patient decisions about knee-replacement surgery. *MDM.* 2014. DOI: 10.1177/0272989X14527796.

4. Rossi, P., & Anderson, A. The factorial survey approach: an introduction. In Rossi, Peter & Nock, Steven (Eds.) *Measuring Social Judgments: The Factorial Survey Approach*. Beverly Hills: Sage. 1982. ISBN 10: 080391816X

5. Kanninen, B. (Ed.) *Valuing Environmental Amenities Using Stated Choice Studies*. Springer. 2007. http://www.springer.com/economics/environmental/book/978-1-4020-4064-1.

6. Mitchell, R., & Carson, R. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington, DC: Resources for the Future. 1989. ISBN-10: 0915707322

7. International Patient Decision Aid Standards Collaboration (IPDAS). (2012). http://ipdas.ohri.ca/index.html International Patient Decision Aid Standards Collaboration (IPDAS). (2012). http://ipdas.ohri.ca/index.html.

8. Tubeuf, S. Edlin, R., Shourie, S., Cheater F., Bekker, H., & Jackson, C.Cost effectiveness of a web-based decision aid for parents deciding about MMR vaccination: a three-arm cluster randomised controlled trial in primary care. BJGP. 64(625)e493-e499 2014,;DOI: 10.3399/bjgp14X680977.

9. Arber, S., McKinlay, J. Adams, A., Marceau, L., Link, C., & O'Donnell, A. Patient characteristics and inequalities in doctors' diagnostic and management strategies relating to CHD: a video-simulation experiment. Soc Sci Med*, 62*(1), 103–15. 2006 PMID: 16002197

10. Dillman, D. , Smyth, J., & Christian, L. *Mail and Internet Surveys: The Tailored Design Method*. Edition 3. New York: Wiley. 2008. 13: 978-0471698685

11. Couper M.P. (2014). The future of modes of data collection. Public Opin Q. Vol. *75*, No. 5, 2011, pp. 889–908. (2014). doi: 10.1093/poq/nfr046

12. Fuchs, M., & Funke, F. Video Web Survey - Results of an experimental comparison with a text-based Web survey. In Trotman, M., Burrell, T., Gerrard, L., Anderton, K, Basi, G., Couper, M., Morris, K., Birks, D., Johnson, A.J., Baker, R., Rigg, M., Taylor, S., & Westlake, A. (Eds.) The Challenges of a Changing World: Association for Survey Computing. 2007. http://www.asc.org.uk/publications/proceedings/ASC2007Proceedings.pdf#page=74
13. Baddeley, A. *Human Memory: Theory and practice*. Needham Heights, Massachusetts: Allyn & Bacon. 1990. ISBN-13: 978-0205279487.
14. Groves, R. *Survey Errors and Survey Costs*. New York: Wiley. 2004. ISBN-13: 978-0471678519
15. De Leeuw, E.D. , Hox, J.J., & Huisman, M. (2003). **Prevention and treatment of item nonresponse**. J Off Stat, 19, 2, 153–176. 2003. http://www.joophox.net/publist/jos03.pdf
16. Bassili, J.N., & Scott, B.S. Response latency as a signal to question problems in survey research, *Public Opin Q.* 60, 390–399. 1996. DOI: 10.1086/297760
17. McNair, J. 2009. http://ezinearticles.com/?What-is-the-Average-Reading-Speed-and-the-Best-Rate-of-Reading?&id=2298503. Last retrieved 03-11-2014.
18. Tourangeau, T., Rips, L.J., & Rasinski, K. *The Psychology of Survey Response*. NY: Cambridge University Press. 2000. ISBN-13: 978-0521576291

Table 1 Descriptive statistics				
Sample	**Assigned text**	**Text fallback**	**Video**	**P-value***
Variable	Mean	Mean	Mean	
Gender				
Male	0.43	0.42	0.44	0.447
Age				
50-59	0.50	0.45	0.51	0.021
60-69	0.34	0.29	0.35	0.049
70 or older	0.16	0.26	0.14	0.000
Income				
below 25,000$	0.23	0.32	0.19	0.000
≥ 25,000$ and < 50,000$	0.27	0.29	0.27	0.397
≥ 50,000$ and < 75,000$	0.16	0.16	0.17	0.922
above 75,000$	0.33	0.23	0.37	0.000
Living Status				
Married or living with a partner	0.58	0.52	0.61	0.004
Employment Status				
Retired	0.34	0.43	0.33	0.000
Working	0.45	0.34	0.48	0.000
Unemployed, disabled and other	0.20	0.23	0.20	0.187
Education				
High school or less	0.24	0.27	0.22	0.080
At most Bachelor's Degree	0.59	0.58	0.61	0.464
Post graduate	0.17	0.15	0.17	0.301
Ethnicity				
Non-Hispanic white	0.83	0.78	0.86	0.012
Respondents with chronic knee pain				
	0.42	0.44	0.41	0.441
Respondents with knee osteoarthritis				
	0.21	0.23	0.21	0.305
Friends/Relatives with knee osteoarthritis				
	0.66	0.67	0.64	0.678
Friends/Relatives with full knee-replacement surgery				
	0.53	0.54	0.51	0.639
Number of observations	339	358	919	

*P-value for the difference in means of the respective variable between the text fallback and video groups

Table 2 Distribution of total time in seconds spent on the survey				
Sample	**Full**	**Assigned text**	**Text fallback**	**Video**
Percentiles				
25th	689	481	665	802
50th	905	637	877	987
75th	1225	988	1317	1267

Table 3 Distribution of time in seconds spent on vignette components

Sample	Full	Assigned text	Text fallback	Video
Total time in seconds spent on:				
Vignette person's situation				
Percentiles				
25th	33	23	23	62
50th	61	32	34	70
75th	74	46	48	81
Surgeon's recommendation				
Percentiles				
25th	16	11	11	31
50th	31	16	17	39
75th	41	24	26	44
Second opinion				
Percentiles				
25th	14	10	10	29
50th	27	14	14	32
75th	34	19	20	37
Outcome forecast				
Percentiles				
25th	10	10	10	11
50th	22	19	20	25
75th	36	31	34	38
First patient testimonial				
Percentiles				
25th	21	16	14	40
50th	39	22	23	48
75th	51	30	31	56
Second patient testimonial				
Percentiles				
25th	16	10	11	37
50th	34	15	16	42
75th	43	22	22	47

Table 4 Distribution of time in seconds spent on decision screen				
Sample	**Full**	**Assigned text**	**Text fallback**	**Video**
First vignette observation				
Percentiles				
25th	5	5	6	5
50th	7	7	8	7
75th	10	10	12	9
Second vignette observation				
Percentiles				
25th	5	4	5	5
50th	6	6	6	6
75th	8	8	10	8
Third vignette observation				
Percentiles				
25th	4	4	5	4
50th	6	6	6	5
75th	8	8	9	7

Table 5 Speeding for vignette dimensions (percentages)				
Sample	**Full**	**Assigned text**	**Text fallback**	**Video**
Percentage of speeding for:				
Vignette person's situation	5.47	6.89	9.03	3.56
Surgeon's recommendation	4.38	6.20	6.98	2.69
Second opinion	2.33	4.63	3.82	0.91
Outcome forecast	5.14	5.41	5.59	4.86
First patient testimonial	3.53	3.83	6.70	2.18
Second patient testimonial	1.98	3.54	4.19	0.55