Collaborative Patient Portals: Computer-based Agents and Patients' Understanding of Numeric Health Information

PI: Dan Morrow

Co-Investigators: Mark Hasegawa-Johnson, Thomas Huang, and William Schuh

University of Illinois Urbana-Champaign

Project Officer: CDR Derrick Wyatt

Project Period 05/01/2014 - 04/30/2017

Agency for Healthcare Research & Quality (R21 HS22948)

FINAL PROJECT REPORT

Abstract

Purpose

Our goal is to improve understanding and use of information in portals to Electronic Health Record (EHR) systems by older adults with diverse numeracy and literacy abilities.

Scope

Portals are underutilized in part because patients struggle to understand numeric health information such as clinical test results. To understand this information, patients must create gist representations capturing the 'bottom-line' for their health. Older adults, the most frequent consumers of health information, are challenged because of limited numeracy and because portals often present numbers with little context.

Methods

We designed portal messages to improve older adult memory for test results by enhancing context (compared to standard format) with verbal (test scores labeled to indicate degree of risk), graphic (scores embedded in number lines) and Computer agent (CA: information spoken by the CA, with verbal and nonverbal cues) formats. In Study 1, a video provider served as proxy for the CA. In Study 2 the CA was compared to the video.

Results

In Study 1, older adults' gist memory for test results was improved by enhanced compared to standard message formats, with some evidence that the video format was most effective. While enhanced formats did not directly improve affective response, risk perception, or behavioral intentions, path analyses showed that gist memory and affective response were associated with risk perception and behavioral intentions. In Study 2, the CA was as effective as the video for supporting gist memory and affective response. These findings suggest the value of CA for conveying numeric information in portals.

Key Words

Computer agent, patient portals, gist memory, self-care

Purpose and Scope

We focused on helping older adults understand numeric information such as their clinical test results because this information is often presented in patient portals, and patients, especially older adults, struggle to make sense of this information. Test results (and other portal-based numeric information) are often presented as a table of numbers with little context (e.g., information about the scale for the scores, or relation of the scores to risk for illness). Clinicians traditionally help patients understand this information by using verbal and nonverbal strategies (e.g., facial expressions, tone of voice) that contextualize the information. However, with decreasing patient contact time, clinicians are challenged to provide this support. HIT has the potential to support patient comprehension by providing access to information outside the constraints of brief clinical encounters. However, portals may not be effective because they eliminate the clinical context that supports patient comprehension.

We developed enhanced portal message formats that help older adults understand numeric health information. According to fuzzy trace memory theory, people understand and remember numeric information at multiple levels. The verbatim level captures the literal or "surface form" of information (precise numeric values). However, making sense of and using numbers also requires interpreting the information in terms of goals and knowledge to create gist-based representations that capture the bottom-line implications for health. Gist is often organized around qualitative dimensions that are affective and evaluative (e.g., good/bad). For instance, gist representations of cholesterol test results may capture ordinal values of risk for heart disease (lower/borderline/higher). Older adults may focus on gist because it is often easier to construct compared to verbatim representations and because they focus on affective information.

We investigated whether message formats designed to enhance the context of standard portal messages about cholesterol and diabetes screening test results improved gist comprehension and risk perception, and boost attitudes toward and intention to perform self-care behaviors. More specifically, we investigated verbally, graphically, and video-enhanced formats that provide increasing levels of support for gist-level understanding of risk (for more information, see Morrow, Hasegawa-Johnson, Huang, Schuh, Azevedo, Gu, Zhang, Roy, & Garcia-Retamero, 2017). The verbally enhanced format provided verbal cues often used in face-to-face communication to help patients understand the risk associated with the test scores. Labels for evaluative categories (low, borderline, or high risk) were added to the message to provide context for interpreting the specific scores and promote emotional processing of the quantitative information. The graphically enhanced format provided graphical as well as verbal context, with the test scores embedded in number lines for each score, with color coding reinforcing the verbal labels of each region of risk on the scale. Emphasizing the order of the regions from lower to higher risk (or higher to lower risk for HDL) should support ordinal gist understanding organized around evaluative and affective dimensions. In the most enhanced message condition, the same graphic display (with the labeled test scores) is accompanied by a video of a physician discussing the results and their implications for risk. The physician provides high-level commentary, using nonverbal cues (prosody, facial expressions) and verbal cues (the same risk category labels as in the other enhanced formats) to signal information relevance and guide affective interpretation. This multimedia video format may be most effective because the verbal and nonverbal cues reinforce each other. In our first message evaluation study, the video of the actual physician was compared to the other enhanced message formats, as well as the standard format. In the second evaluation study, a computer agent developed from the video physician was compared to the video format.

Developing the Computer Agent. The video message format would be difficult to implement in actual portals because different videos would be required for each patient and set of test results. Therefore, a focus of our project was to develop a computer-based agent (CA), or 'virtual provider' to deliver health information in portal environments. Ideally, this CA emulates best practices for face-to-face communication but is also generative, able to deliver a wide range of health information to diverse patients. In our portal format evaluation studies, this CA-enhanced format was identical to the video condition, except the video is replaced by a CA version of the video recorded physician (for more detail see Morrow et al., 2017).

The CA builds on progress in several fields developing CAs for human-computer systems. In general, CAs with realistic facial expressions, gestures, and other relational cues improve learning in tutoring systems compared to text-based systems, in part by engendering social responses. CAs are often used and evaluated in education settings, but their use and evaluation in health care are less frequent. The CA was developed in three phases: The CA's visual appearance was generated from a frontal face image of the physician; the audio (synthesized voice) of the CA was generated using a speech synthesizer; and the visual and audio components were then synchronized.

*Developing the CA visual appearance***.** A 2D frontal face image of the physician in neutral expression was used to generate the CA's 3D appearance, using an algorithm that localizes the feature points on the face and fits a 3D morphable generic model to the image. Once generated, the 3D CA was animated to simulate a real person talking, so that the CA makes appropriate lip movements that match spoken phonemes, as well as facial expressions in order to match the input text with emotion markers. In our implementation, the CA is controlled by Facial Action Parameters (FAP). For example, the "stretch lip corner" FAP specifies how much the lip corner of the 3D CA is stretched to represent smiling or certain lip shapes during speech. Different FAPs are combined to control the whole face to match the CA appearance with the speech and emotion context.

The lip movements are mainly affected by speech content and emotion. To integrate these two factors, we define a template for each lip movement under each expression separately. During the synthesis process, all lip movements are predefined and the best fit template selected. Finally, we construct keyframes, where the CA has expressions and lip shapes matching the target viseme and emotion marker, by selecting the appropriate templates and interpolating between the keyframes to generate the talking head.

*Developing the CA synthetic voice***.** A physician from our partner health organization first recorded 8 hours of speech, including the scripts for the video messages for the message evaluation experiments as well as other text needed to develop the CA. The resulting audiovisual recordings of the physician's face were segmented using two automatic speech recognition toolkits: HTK and Kaldi. Therefore, the baseline audio speech synthesizers were developed using first a unit selection concatenation of audiovisual segments from this recorded corpus. The speech in the corpus was segmented into phone units and a synthetic utterance produced by selecting and concatenating these units that constitute the target utterances, maximizing continuity and naturalness. To capture emotional nuance, each sentence was assigned positive, neutral and negative labels, based on the valence of the medical test results in that sentence. To synthesize an utterance with a specific emotional valence, the synthesizer placed a preference on the phone units with the matched valence label. However, pilot testing found that participants could not determine the valence of the utterances signaled exclusively by the origin of the audio clips (in a concatenative synthesizer), without corresponding changes to the utterance content. It is possible that the emotional nuance was overwhelmed by degraded naturalness of the utterance, or unit-selection was unable to capture emotional nuances given the limited corpus.

We next focused on pinpoint modification of F0 contours correlated with specific syntactic and semantic information in the text. Differential linear regressions were used to learn the difference between F0 (pitch) of any given word with versus without pitch accent, and texts were synthesized with pitch accents added in appropriate positions. Next, this approach was generalized to learn differences in pitch contour between the speech of two a monotone and a dynamic speaker. The hidden F0 targets underlying observed F0 measurements in each syllable were estimated and learned using Xu's "Parallel Encoding of Target Approximation" (PENTA) model. Mapping between PENTA targets of the monotone and dynamic speakers was learned and the speech of the monotone speaker was processed to have a pitch contour better matching that of the dynamic speaker. We leveraged this approach to guide post-processing of utterances produced by the Vocalware speech synthesizer for the CA.

Synchronizing Video and Audio. Finally, the synthesized face animation and speech audio components were synchronized. The audio and video were aligned via keyframes, where each keyframe corresponds to one viseme. The input text (the CA messages) were first converted into visemes (phoneme for audio synthesis). For each viseme, a keyframe was generated with a visual template (defining the lip shape and expression) associated with it. For the video synthesis, the gaps between each keyframe are filled by interpolation to generate a smooth lip shape transition. For audio synthesis, the selected units are simply concatenated to generate the speech audio. To summarize, the CA system takes a 2D picture and text script with emotion markers as input. The emotion markers indicate the desired emotion for each sentence. Finally, the CA system generates a 3D talking head as output.

Methods

Developing a methodology to evaluate portal messages. To evaluate the portal message formats, we used a multi-method approach that combined quantitative and qualitative measures, as well as experimental and individual difference methods (Morrow et al., 2017). As part of developing the messages, pilot studies were conducted to finalize the format design, leading to formal scenario-based experimental studies that evaluated whether enhanced formats improved memory for and use of self-care information compared to the standard format. Study 1 focused on the video-enhanced format, while Study 2 compared CA-enhanced to video-enhanced formats. The scenarios for the formal studies were developed in collaboration with two physicians from our partner health care organization. Each scenario contained a fictitious patient profile and a message describing cholesterol or HbA1c diabetes screening test results for that patient. Profiles were included because risk level depends on patient-specific risk factors (e.g., family history of heart disease) as well as specific test results. Diabetes as well as cholesterol test results were evaluated in order to generalize results across type of test. The diabetes messages described a percent score on one test (plasma glucose concentration, A1C), while the cholesterol messages describe more complex patterns of scores on multiple tests (total cholesterol, triglycerides, high density lipoproteins (HDL), and low density lipoproteins (LDL)) that suggest lower, borderline, or higher risk for cardio-vascular illness. To help patients understand the overall risk associated with each message, the message ended with a summary of the overall risk associated with the test scores. For the enhanced messages, the summary conveyed the gist risk category.

The messages describing the test results in the scenarios were presented in one of the four formats described above, and participants were randomly assigned to one of the conditions.

After viewing each scenario, participants responded to a series of questions that measured verbatim and gist memory for the test results, perceived risk, which may reflect motivation to act so as to mitigate that risk, affective reactions to the risk information in the messages, attitude toward taking medications to treat illness suggested by the test results, and intention to perform self-care activities related to reducing risk of disease (following medication, diet, and exercise recommendations).

Pilot Study: Evaluating video messages. Before conducting the formal evaluation studies, a pilot study was conducted to investigate whether the video messages themselves (i.e., without the graphic displays) were easy to understand and would prompt appropriate affective responses (Azevedo, Morrow, Hasegawa-Johnson, Gu, Soberal, Huang, Schuh, & Garcia-Retamero, 2015). We wanted to validate the video messages because the physician in the videos served as the template for developing the CA to be used in the CA-enhanced message condition. We were most interested in whether older adults could identify level of risk associated with the test results based solely on the physician’s verbal and nonverbal cues in the video messages.

Twelve older adults viewed the video messages describing the cholesterol and diabetes clinical test results. After each message, participants pretended to be the patient and responded to questions that probed gist memory, affective reaction to the test results, and satisfaction with the message. We also asked whether the way the information was delivered by the physician matched the information provided. Participants considered both the cholesterol and diabetes messages to be very appropriate and the information to be very useful. They were also generally able to understand the gist of the messages, and their affective responses were appropriate to the level of risk associated with the test scores: As risk increased, positive affect decreased and negative affect increased. These results suggest that participants developed a gist representation organized in terms of affective as well as cognitive dimensions.

Message Evaluation Study 1. Encouraged by the pilot study, we conducted the first formal message evaluation study investigated whether enhancing the context for standard portal messages about test results improved older adult gist comprehension of risk associated with the results and increased attitudes toward and intention to perform self-care behaviors that can mitigate risk (Morrow, Azevedo, Garcia-Retamero, Hasegawa-Johnson, Huang, Schuh, Gu, & Zhang, 2017). 144 older adults viewed the scenarios with messages that described results from cholesterol or diabetes screening tests that indicated lower, borderline or higher risk levels. These messages were conveyed by standard format (a table of numerical test scores) or by one of the enhanced formats (verbal, graphic, video).

We found that enhancing the context of numeric information improved older adults' memory for and response to clinical test results in the portal messages compared to the table-based format often used in patient portals to EHR systems. More specifically, enhanced messages improved gist but not verbatim memory for test results, consistent with fuzzy trace theory. This result also suggests we were successful in designing portal messages that highlight gist-based risk information. In addition, we found some evidence that the video-enhanced and verbally enhanced formats were more effective than the graphic format, with some evidence that the video format was most effective (overall memory for component scores in the messages; global gist memory).

An unexpected finding was the tendency for the older adults in the study to estimate higher levels of risk for the 'good news'/low risk messages in the graphically enhanced condition compared to the other formats. This result is consistent with findings in the risk perception literature suggesting that some types of graphic displays can encourage risk avoidance (e.g., perhaps by emphasizing the full range of risk values for a score, including the high risk regions). This may be especially the case for older adults because of an age-related decline in controlled attention processes that inhibit irrelevant information, so that they misremember low risk scores as higher risk. Evidence for this explanation comes from a recently conducted study where we found suggest that younger adults who viewed the same graphically enhanced messages estimated lower risk levels than the older adults did for the 'good news' test result messages (Malik et al., 2017). Interestingly, younger adults also reported lower intention to perform self-care in this condition compared to the older adults, perhaps reflecting their inflated lower risk perception. Older adults also responded with higher levels of positive affect to the messages than younger adults, consistent with the positivity effect and more generally, evidence for age-related focus on emotional self-regulation.

Notably, older adults did not overestimate risk (for low risk messages) in the video-enhanced condition even though it contained the same graphic. It is possible that the physician commentary in the video condition countered this effect by emphasizing how the test scores are associated with specific regions of the scale and reducing the impact of the salient higher risk regions on the number line, suggesting the value of multimedia for conveying risk information to older adults in patient portals.

Finally, the level of risk associated with the test results had a robust effect on almost all measures in the study. However, risk level had different effects on message memory compared to participants' responses to this information: While participants recalled borderline results less accurately than either lower or higher risk, their affect, risk perception, attitude toward, and intent to perform health behaviors increased monotonically with risk level. They may have remembered borderline results less accurately because the component scores varied in level of risk indicated, which complicated the process of integrating the information into an overall estimate of risk for these messages. In addition, participants tended to remember the gist of borderline risk messages as higher than indicated in the messages, which may have had the effect of increasing perceived risk, affective response, behavioral attitude, and intent in this condition. This finding, along with the finding that older adults' overestimation of risk for low risk messages in the graphic condition was associated with heightened risk perception and behavioral intention, is consistent with behavioral change theories and suggest the value of gist-based formats for influencing motivation to perform self-care. Further evidence was provided by a multi-variate path analysis of the results (collapsed over message format to increase sample size), which showed that gist memory for and affective response to the messages predicted risk perception and attitude toward and intention to perform self-care behaviors that may mitigate the risk (Azevedo et al., 2017).

Pilot Study 2: Evaluating CA Voice. The second message evaluation study focused on evaluating the CA format in comparison to the video format. Before conducting this study, a pilot study was conducted to examine the impact of synthesized versus natural voice on older adults' memory for clinical test results as part of refining the CA (for more detail see Morrow et al., 2017).

Twenty-four older adults listened to a subset of the cholesterol messages that described results indicating lower, borderline, or higher risk. The messages were delivered either using the physician's voice from the video messages (natural voice condition) or the Vocalware speech synthesizer (synthesized voice). In both voice conditions, participants viewed the test results presented in standard format as they listened to the message. After each message, gist memory, perceived risk, and message satisfaction were measured. The participants then listened to natural and synthetic voice versions of the same message and indicated which voice (natural or synthesized) they preferred and why. We expected minimal differences in gist memory for the two types of voice messages (although there is some evidence that older adults better remember natural versus synthesized voice messages compared to younger adults), even if participants prefer the natural voice messages.

Gist memory was accurate for both conditions, with no difference between the two (Natural voice = 88%, Synthesized = 90% correct). Perceived risk and message satisfaction were also similar for the two conditions. Moreover, when participants directly compared natural and synthetic versions of the same message, there was not a reliable difference in preference for the two voices (natural=58%; synthetic=42%, binomial $p > .50$). Those preferring the natural voice mentioned that this voice sounded more human, conversational, and empathetic compared to the synthetic voice. On the other hand, those preferring the synthetic voice mentioned that this voice was clearer and attracted attention. Therefore, even though there was a numeric trend in favor of the natural voice in a direct comparison with the synthetic voice, participants were satisfied with and accurately remembered and responded to the risk information in both types of messages. Based on these findings, as well as the fact that synthesized voice would be more feasible than recorded speech for large-scale implementation in patient portals, we decided to use the synthesized voice in the message evaluation experiment.

Message Evaluation Study 2. This study evaluated the CA, using the same scenario-based materials and measures developed for the first message evaluation study (Azevedo et al., 2017). Performance in the CA condition was compared to performance from the video-enhanced (video recorded provider explaining test results in the portal message) and from the standard portal conditions in Study 1. We expected the CA-enhanced portal messages and the video-enhanced messages to be more effective than the standard portal format for conveying the clinical test result information because the multimedia cues promote gist-based memory for risk and appropriate affective response. The CA's and video provider's nonverbal cues should boost intention to act because they engender affective and other social responses. We also expected no significant differences between the CA and video conditions. Even if the CA is not more effective than the video, there are compelling reasons to implement the CA in portal (and other) environments because the CA will be a generative communication system that creates new messages for a wide range of health applications and patients.

Thirty-six older adults viewed the same scenarios as in the previous study, but with CA-enhanced rather than video-enhanced messages. Like the video messages, the CA-enhanced messages included a graphic format in which risk level of the test scores were conveyed by number lines with color coding used to convey level of risk and appropriate affective response to this information. Gist memory for the risk information in the messages was high and similar for the CA and video conditions (Video = 87.5%, CA = 84.7% accuracy; $p = 0.5$; Cohen's $d < 0.2$). Moreover, both of the enhanced groups better remembered risk information compared to those in the standard format condition, both formats improved comprehension (e.g., CA = 84.7%, Standard = 67.6% accuracy; $p < .01$; Cohen's $d < 0.6$). Participants were also equally satisfied with the two enhanced message and more so than the standard format. However, when participants in the CA group were presented both the video and the CA at the end of the session, they significantly preferred to receive their test results from the video of a real person (Video = 81.9%, CA = 18.1% preference; $p < .01$, Cohen's $d = 2.0$). Participants preferred the videos because the provider was seen as more relatable, personable, conversational, natural sounding, or as explaining more (despite both formats delivering identical content).

Results

Findings from our project support our general approach to improving older adult understanding and use of numeric information provided in patient portal environments. The first evaluation study showed that enhancing the context of numeric information improved older adults' memory for and response to clinical test results described by portal messages compared to the table-based format often used in patient portals to EHR systems. Guided by our framework that integrates health literacy, fuzzy trace memory, and behavior change theories, we designed portal messages that highlight gist-based risk information associated with test results. In addition, we found some evidence that the video-enhanced and verbally enhanced formats were more effective than the graphic format, with some evidence that the video format was most effective (overall memory for component scores in the messages; global gist memory). We also provided some evidence that validated our approach to developing a CA-based format for patient portals.

The second evaluation study showed that both the video recording and the CA effectively communicated facts about cholesterol, and inspired patients' satisfaction. However, while the CA was intelligible, there is potential to improve the CA appearance and other features in order to make it more preferred compared to an actual person. However, we note that there is debate about the appropriate level of realism for CAs (animated vs static) that produces trust and 'social stance' that primes learning. Increasing realism until the CA appears almost, but not quite, human can produce negative responses (e.g., 'uncanny valley' effect). The photo-realistic CA in our project, may fall into the valley, so that less realistic versions of the CA will be more effective when compared to a human recording.

Another challenge identified by our findings is to design portal messages that engage patients so that they are motivated to act on, as well as understand and remember the messages. Our enhanced formats improved gist memory for risk information, which in turn was associated with affective response, risk perception, and intention to perform self-care behaviors such as change diet and exercise. However, these formats only indirectly influenced intention. We are now exploring a range of CAs that vary in gender, age, and realism in order to investigate whether more expressive CAs (or having choice over the CA) has a greater impact on older adult motivation related to self-care. We also hope to investigate other types of communication in portals, such as patient stories about self-care, which have been found to influence patient attitude and performance of health behaviors.

The CA should support these intentions by promoting affectively organized gist representations that motivate action. While portals may one day revolutionize patient-centered care by increasing access to information, many older do not use them. We will help make portals more cognitively accessible by making information more easily understood and actionable.

Publications

Azevedo, R.F.L., Garcia-Retamero, R. Hasegawa-Johnson, M., Huang, T., Schuh, W., Gu, K., Zhang, Y., Yang, X., & Morrow, D.G. (2017, Aug). A multidisciplinary approach to improving patient portal messages for patients with diverse abilities. *American Psychological Association's Annual Meeting*, Washington DC.

Azevedo R.F.L., Gu, K., Zhang, Y., Sadauscas, V., Sakakini, T., Morrow, D.G., Hasegawa-Johnson, M., Soberal, D., Huang, T., Bhat, S., Willemsen-Dunlap, A., Halpin, D., Graumlich, J., & Schuh, W. (2017) Using computer agents to explain clinical test results. *Proceedings of the AMIA 2017 Annual Symposium*. American Medical Informatics Association,

Azevedo, R.F.L., Morrow, D, Hasegawa-Johnson, M., Gu, K., Soberal, D., Huang, T., Schuh, W., & Garcia-Retamero, R. (2015). Improving patient comprehension of numeric health information. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting*. Santa Monica: Human Factors & Ergonomics Society.

Azevedo, R. & Morrow, D.G. (2017). Improving older adult comprehension and use of EMR portal-based information. In R. Pak & A. McLaughlin (Eds.), *Aging, Technology, and Health*. Elsevier. In press.

Malik, T., Azevedo, R.F.L., Morrow, D.G., Hasegawa-Johnson, M., Huang, T., Schuh, W., Gu, K., Garcia-Retamero, R. (2017). Age differences in responding to graphically conveyed health risk information. *Proceedings of the Human Factors and Ergonomics Society 61st Annual Meeting*. Santa Monica: Human Factors & Ergonomics Society.

Morrow, D.G. Azevedo, R.F.L., Garcia-Retamero, R. Hasegawa-Johnson, M., Huang, T., Schuh, W., Gu, K., & Zhang, Y. (2017). *Contextualizing Numeric Clinical Test Results for Gist Comprehension: Implications for EHR Patient Portals***.** In preparation.

Morrow, D.G. Hasegawa-Johnson, M., Huang, T., Schuh, W., Azevedo, R.F.L., Gu, K., Zhang, Y., Roy B., & Garcia-Retamero, R. (2017). A multidisciplinary approach to designing and evaluating electronic medical record portal messages that support patient self-care. *Journal of Biomedical Informatics. 69*, 63–74. PMID: 28347856

Morrow, D.G., Hasegawa-Johnson, M., Azevedo, R.F.L., Chong, C., Roy, B., Gu, K., Huang, T., Schuh, W., & Garcia-Retamero, R. (2015, Oct). Can computer agents improve comprehension of patient portal information? *The 2015 AHRQ Research Conference*, Crystal City, VA.