



HHS Public Access

Author manuscript

Digit Cult Educ. Author manuscript; available in PMC 2018 August 17.

Published in final edited form as:
Digit Cult Educ. 2018 ; 10: 22–48.

“I DIDN’T TELL YOU SOONER BECAUSE I DIDN’T KNOW HOW TO HANDLE IT MYSELF.” DEVELOPING A VIRTUAL REALITY PROGRAM TO SUPPORT HIV-STATUS DISCLOSURE DECISIONS

Kathryn E. Muessig, PhD [Assistant Professor],

Department of Health Behavior, CB #7440, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7440

Kelly A. Knudtson, MPH [Project Manager],

Institute for Global Health and Infectious Diseases, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Karina Soni [Project Coordinator],

Institute for Global Health and Infectious Diseases, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Margo Adams Larsen, PhD [Director of Research],

Virtually Better Inc., 2440 Lawrenceville Hwy, Suite 200, Decatur, Georgia 30033

David Traum, PhD [Research Assistant Professor],

USC Viterbi School of Engineering Computer Science Department, University of Southern California (USC), Director for Natural Language Research, USC Institute for Creative Technologies (ICT), 12015 Waterfront Drive, Playa Vista, CA 90094-2536

Willa Dong, MSPH,

Department of Health Behavior, Institute for Global Health and Infectious Diseases, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Donaldson F. Conserve, PhD [Assistant Professor],

Department of Health Promotion, Education, and Behavior, Arnold School of Public Health, University of South Carolina, 915 Greene St, Columbia, SC 29208

Anton Leuski, PhD [Research Assistant Professor],

USC Viterbi School of Engineering Computer Science Department, University of Southern California (USC), USC Institute for Creative Technologies (ICT), 12015 Waterfront Drive, Playa Vista, CA 90094-2536

Ron Artstein [Research Assistant Professor], and

USC Viterbi School of Engineering Computer Science Department, University of Southern California (USC), USC Institute for Creative Technologies (ICT), 12015 Waterfront Drive, Playa Vista, CA 90094-2536

Corresponding author: Kathryn E. Muessig, PhD, Assistant Professor, Department of Health Behavior CB #7440, Gillings School of Global Public Health The University of North Carolina at Chapel Hill Chapel Hill, NC 27599-7440 919-962-5059 (phone) 919-966-2921 (fax) kmuessig@med.unc.edu.

Lisa B. Hightow-Weidman, MD, MPH [Associate Professor of Medicine and Health Behavior] [Director, Behavior and Technology Lab]

Institute for Global Health and Infectious Diseases, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Abstract

HIV status disclosure is associated with increased social support and protective behaviors against HIV transmission. Yet disclosure poses significant challenges in the face of persistent societal stigma. Few interventions focus on decision-making self-efficacy, and communication skills to support disclosing HIV status to an intimate partner. Virtual reality (VR) and artificial intelligence (AI) technologies offer powerful tools to address this gap. Informed by Social Cognitive Theory, we created the Tough Talks VR program for HIV-positive young men who have sex with men (YMSM) to practice status disclosure safely and confidentially. Fifty-eight YMSM (ages 18 — 30, 88% HIV-positive) contributed 132 disclosure dialogues to develop the prototype through focus groups, usability testing, and a technical pilot. The prototype includes three disclosure scenarios (neutral, sympathetic, and negative response) and a database of 125 virtual character utterances. Participants select a VR scenario and realistic virtual character with whom to practice. In a pilot test of the fully automated neutral response scenario, the AI system responded appropriately to 71% of participant utterances. Most pilot study participants agreed Tough Talks was easy to use (9/11) and that they would like to use the system frequently (9/11). Tough Talks demonstrates that VR can be used to practice HIV status disclosure and lessons learned from program development offer insights for the use of AI systems for other areas of health and education.

Keywords

virtual reality; HIV; disclosure; men who have sex with men; artificial intelligence

Introduction

In 2014, 67% of all new HIV diagnoses in the United States (US) were among men who have sex with men (MSM), with young MSM (YMSM), ages 13 to 34, facing the highest rates of new infections (U.S. Centers for Disease Control and Prevention, 2015). Among MSM, HIV status disclosure may facilitate lower sexual risk behaviors and greater medication adherence (Hightow-Weidman et al., 2013; Morin et al., 2005; Stirratt et al., 2006) thus contributing to reduced HIV transmission (O’Connell, Reed, & Serovich, 2015; Pinkerton, Holtgräve, & Galletly, 2008). Yet disclosing a positive serostatus poses challenges for many YMSM living with HIV and is complicated by a range of factors — most prominently stigma, discrimination, and confidentiality concerns (J. Bird, Eversman, & Voisin, 2017; Overstreet, Eamshaw, Kalichman, & Quinn, 2013; Przybyla et al., 2013; Ronn, White, Hughes, & Ward, 2014; Serovich & Mosack, 2003); indeed, half or less report disclosing to steady partners or other sexual partners (Cook, Valera, & Wilson, 2015; Grov, Rendina, Moody, Ventuneac, & Parsons, 2015; Hightow-Weidman et al., 2013). Youth living with HIV face additional barriers to disclosure including fear of rejection and lack of communication skills for disclosing (Thoth, Tucker, Leahy, & Stewart, 2014). For HIV-positive MSM, condomless sex with nondisclosure is associated with worse HIV medication

adherence, less viral suppression, lower self-efficacy to disclose, and less frequent use of risk-reduction communication (Kalichman, Kalichman, Cherry, & Grebler, 2016).

Despite the importance and difficulty of HIV status disclosure, interventions for MSM are limited and those tailored for racial and ethnic minority YMSM are nonexistent (Conserve, Groves, & Maman, 2015). Thus while public health messaging and health care providers increasingly promote HIV status disclosure (Persson & Richards, 2008; U. S. Centers for Disease Control and Prevention, 2014), there is limited recognition in this messaging of the challenges inherent in disclosure and minimal resources to support YMSM in disclosure decision making and execution.

A small number of HIV status disclosure interventions for MSM have employed electronic health delivery platforms (eHealth) with mixed success to provide education and tailored feedback, foster self-awareness of sexual-risk, and facilitate creating disclosure plans and practicing disclosure scenarios (Bachmann et al., 2013; Chiasson, Shaw, Humberstone, Hirshfield, & Hartel, 2009; Milam et al., 2016; Serovich, Reed, Grafsky, & Andrist, 2009). eHealth approaches may be particularly relevant for supporting disclosure because digital platforms are already broadly used by YMSM to find sexual partners and can facilitate communication about HIV status (Broaddus et al., 2015; Grov, Agyemang, Ventuneac, & Breslow, 2013). Two studies that used computer-based platforms to deliver tailored disclosure support messaging (Bachmann et al., 2013) and skills-building (Serovich et al., 2009) found no improvements in disclosure. In contrast, two studies — one utilizing a monthly internet self-report risk survey (Milam et al., 2016) and another using a brief video drama delivered online (Chiasson et al., 2009) found positive increases in HIV status disclosure. These mixed results from eHealth disclosure intervention designs are encouraging, however the complexities and challenges of HIV status disclosure for diverse YMSM suggest that additional digital approaches should be explored.

Virtual reality (VR) technologies may be uniquely able to create highly engaging, effective disclosure interventions. VR offers users a safe space to learn and practice problem-solving in realistic, “low-stakes” environments and subsequently apply these skills in real life (Durlach & Mavor, 1995; Huang, Rauch, & Liaw, 2010; Riva et al., 2007). Learning is enhanced by eliciting emotions and invoking affective experiences through interaction with VR program contents. Public health based VR interventions have achieved behavioral changes in areas ranging from post-traumatic stress disorder therapy, to obesity, diet, and diabetes (Ahn, 2015; Morie, Antonisse, Bouchard, & Chance, 2009; Morie & Chance, 2011). While there are no published VR interventions for HIV status disclosure, an online randomized controlled trial among HIV-negative MSM, ages 18 to 24, demonstrated that interactive virtual environments could significantly reduce sexual shame and subsequent sexual risk behaviors (Christensen et al., 2013). Given evidence of the links between shame, disclosure self-efficacy and HIV status disclosure, we theorized that VR scenarios could be used to improve HIV status disclosure self-efficacy and develop and practice disclosure-relevant skills (e.g. sexual communication, disclosure cost-benefit decision making, and emotional regulation and coping). In this paper, we describe the development and pilot testing of an innovative VR intervention — Tough Talks — that supports serostatus disclosure skills-building and decision-making among YMSM living with HIV.

Developing the Tough Talks Virtual Reality Intervention

Tough Talks was created through a collaboration of the University of North Carolina at Chapel Hill (UNC), the University of Southern California Institute for Creative Technologies (USC-ICT), and Virtually Better Inc. (VBI). This multidisciplinary team included experts in HIV prevention and care, clinical psychology, behavior modification, modeling, simulation, computer science, natural language processing (NLP), VR animation, 3D character modeling, automation technology paradigms, and product development.

Drawing on Social Cognitive Theory (SCT), the program aims to increase self-efficacy for serostatus disclosure through mastery experiences in realistic virtual settings (Bandura, 2011). Individual behavior is uniquely determined through the interaction and mutual influences of behavior, personal/cognitive factors, and environmental influences; behavioral modeling, guided practice, and feedback on performance are expected to result in increased self-efficacy to perform specified behaviors (Bandura, 1986, 2011). SCT posits that people learn through their own experiences as well as by observing the actions of others and the results of those actions. This focus on appraisal and skills building acknowledges that disclosure is typically a selective process that occurs at multiple time points over the course of living with HIV (Chaudoir, Fisher, & Simoni, 2011; Dima, Stutterheim, Lyimo, & de Bruin, 2014). Prior studies on SCT and HIV disclosure have found that low self-efficacy was associated with less disclosure (Abler et al., 2015; Chaudoir et al., 2011; Li, Chen, & Yu, 2016; Serovich, Laschober, Brown, & Kimberly, 2017). Performance accomplishment, or learning through personal experience, is a major source of self-efficacy. Role playing, which can be viewed as a type of performance accomplishment, has been shown to increase self-efficacy (Mausbach, Semple, Strathdee, Zians, & Patterson, 2007). Outcome expectations are crucial for disclosure (Li et al., 2016; Serovich et al., 2017); past experiences impact current disclosure behaviors (Chaudoir et al., 2011). Given the complex social contexts of stigma and potential implications of disclosure in one's life, we sought to create a program that would provide YMSM opportunities to practice disclosing in a safe and confidential environment using a variety of strategies and experiences of different disclosure outcomes and partner responses including acceptance, confusion, lack of HIV knowledge, and rejection.

Informed by SCT, Tough Talks utilizes the technologies of VR and NLP to guide participants through HIV status disclosure conversations. We used an iterative development process to create the first prototype of the Tough Talks program through three phases:

Phase 1: focus groups for developing program requirements and initial development of NLP;

Phase 2: iterative development including three rounds of one-on-one usability testing to refine the Tough Talks program and train the automated artificial intelligence (AI) system; and

Phase 3: pilot-testing to evaluate the program's performance and acceptability. Each of these three phases is described in more detail below. In between the phases, we summarize program developments that resulted from findings in the prior phase.

Table 1 provides an overview of the entire development process, including the research methods used, sample sizes, program components, and study outcomes. Additional details about the software tools used to build Tough Talks are provided in a Technical Appendix.

Study participants (all phases)

Participants were recruited through word of mouth and advertisements at sites (e.g., three high volume HIV clinics, local college campuses) in Raleigh, Durham, Charlotte, and Chapel Hill, North Carolina. Participants were also recruited via banner ads and targeted messages on Craigslist, Facebook, and Grindr. In Phase 1 (focus groups) eligible participants were ages 18–30, assigned male sex at birth and identified as male, reported engaging in anal sex with a male partner within the past 12 months, and English-speaking. Phases 2 (usability testing) and 3 (pilot trial) included the same eligibility criteria for Phase 1 as well as self-reported, behaviorally acquired HIV-positive status.

Interested individuals were screened for eligibility by phone. Those who were eligible were scheduled for an in-person session (focus group, usability session, or pilot-test session) at one of four study venues based on their location and preference. Study staff conducted individual informed consent procedures in-person prior to enrollment. Additional phase-specific methods and analysis procedures are described below under their respective study phase. This study was approved by the Institutional Review Board at UNC.

Phase 1: Focus groups for initial program requirements and development

Phase 1 Methods—Phase 1 aimed to develop the components and content of the VR program that reflected true-to-life YMSM’s HIV status disclosure experiences. We conducted four 90-minute focus groups with HIV-positive (two groups) and HIV-negative or status unknown (two groups) YMSM. Focus groups were conducted in two cities in private rooms within an MSM-focused community based service organization and a community based clinical research study office. Each participant first completed a brief sociodemographic survey. Focus groups were then facilitated by the first author and a trained research assistant both of whom have experience working with YMSM living with HIV (Muessig, Baltierra, Pike, LeGrand, & Hightow-Weidman, 2014; Muessig et al., 2017; Muessig et al., 2013). A semi-structured focus group guide included prompts related to past HIV status disclosure experiences, strategies, barriers, and facilitators. Participants also worked in pairs to write and perform intimate partner HIV status disclosure dialogues. To inform virtual character development, YMSM rated attractiveness of potential partners in photographs and discussed the realism of sample virtual characters. Participants were also asked to evaluate images (photographic and VR) of disclosure conversation locations (e.g., park, bar, cafe, car, home). Lastly, participants discussed the desired functionality of a virtual program coach - a feature that would guide participants through the intervention. Focus groups were digitally recorded and audio recordings were professionally transcribed verbatim (verbalink.com).

Phase 1 Analysis—Data gathered in Phase I were used to build the initial Tough Talks system. Transcripts were inductively analyzed by the UNC team (KK, KM, WD, DC) using

Dedoose (Version 7.1). A codebook was developed including *a priori* and emergent themes to categorize types of disclosure experiences, strategies for managing disclosure situations, types of reactions to disclosure, and specific language used in disclosure conversations. A database of possible conversation phrases (utterance database) was built that the program's virtual characters used to ask questions, respond, and converse with participants. To build the initial utterance database, participant disclosure conversations from the focus group disclosure role play activity and discussions were coded and abstracted verbatim into an Excel spreadsheet. We then expanded the utterance database by adding variations of each utterance and creating new utterances to ensure all focus group themes were captured (e.g., theme: partner's fear of infection; added utterance: "Could I have it too?"). Utterances were developmentally specific, content-appropriate for youth, and culturally sensitive to a variety of racial and ethnic backgrounds and sexual identities. To facilitate the development of the AI program, utterances were designed to convey one thought at a time and limited to 80 characters long, including spaces.

Phase 1 Results—Analyses from Phase 1 resulted in the selection and development of three virtual characters as potential disclosure partners (Figure 1.A) and two settings in which to practice disclosing (Figure 1.B). At the close of Phase 1, the utterance database included 156 virtual character utterances. We created an initial disclosure dialogue policy for these utterances that categorized each utterance by conversation stage (e.g., initiation, pre-disclosure, disclosure, post-disclosure, small talk) and nature (e.g., conversation prompt: "So what did you want to talk about?"; clarification: "What do you mean by that?"). The virtual character utterances covered a range of disclosure reactions identified in the focus groups (e.g., anger, fear, rejection, blame, ignorance, curiosity, confusion, support, concern, sympathy, empathy, acceptance, love). Given time and budget constraints, we focused on developing the three disclosure scenarios that reflected the most common types of disclosure experiences shared in the focus groups for partner type (unexposed primary partner, potentially exposed primary partner post-condomless anal intercourse, and potentially exposed casual partner post-condomless anal intercourse) and partner reaction (e.g., neutral, sympathetic, negative) (Table 2). We included rules in the initial disclosure dialogue policy to guide how utterances could be used within each disclosure scenario. For example, the utterance "I can't believe you're telling me this now." is only applied in the negative partner reaction scenario.

Initial System Development

Following the results of Phase 1, we created an initial clinician-operated version of the VR system. Using a laptop, participants selected from among three virtual characters and two virtual environments. In Phase 1, interaction with the virtual character was text-based. In this version of the system, a study staff member acting as the "Wizard of Oz" (Dahlback, Jonsson, & Ahrenberg, 1993) (wizard) controlled the virtual character's responses via an iPad in order to test whether the set of virtual character utterances and disclosure dialogue policy were sufficient to engage the participant and simulate a realistic disclosure dialogue. During the conversation, the wizard could also create new lines of dialogue if no appropriate preprogrammed utterance existed (Figure 1.C). Disclosure dialogues were captured by the

system's database and subsequently used to refine the utterance database and develop the automated component of the program (Phase 2). These dialogues were also used to “train” the program how to replicate the natural flow of a disclosure dialogue and appropriately respond to new inputs from participants (Artstein, Gandhe, Leuski, & Traum, 2008; Leuski & Traum, 2008, 2010, 2011).

The initial version of the system was fully manual (wizard operated). As the wizard refined the set of utterances used, a first version of an automated system was created. We employed a selective approach to creating automated virtual character responses whereby the system selects from a pre-programmed set of available utterances (the utterance database described above). The system uses a model for selecting utterances that it teams from the wizard's responses. The program's automated selection was then provided to the wizard as a suggested choice, which the wizard could accept (if appropriate for the context of the discussion) or reject and opt to manually choose a more appropriate response. Both the participant and the virtual character were able to “initiate” and “respond” to conversations. In subsequent iterative versions of the system, we implemented policies that constrained some utterances to the proper conversation stage of the dialogue. For example, we constrained utterances coded as reactions to a disclosure statements (e.g., “Could I have it too?”) to the post-disclosure stage of the dialogue. This version of the program was subjected to additional rounds of usability testing and iterative refining as described below in Phase 2.

Phase 2: Usability testing to refine the Tough Talks program and train the automated artificial intelligence (AI) system

Phase 2 had the overall goal of specifying and refining the Tough Talks system, including several important sub-goals. These included:

- narrowing down an essential set of utterances used by the virtual character for voice actor recording and virtual character animation;
- developing a specific policy for how to use the virtual character utterances to carry on a conversation and program this policy to be automated by the AI system;
- evaluating the acceptability and utility of the disclosure dialogues to the participant;
- providing training data and an opportunity to refine the initial AI system; and
- evaluating the performance of the initial AI system and improved system versions.

Details about the technical development of the program features are provided in Technical Appendix A. In this section, we focus on how we engaged participants in the process of evaluating the program, including how we used participants' practice disclosure dialogues to refine the utterance database.

Phase 2 Methods—An iterative development process was applied across three rounds of usability testing (Dumas & Redish, 1999) with YMSM living with HIV. Participants were

recruited, screened, and enrolled following the procedures described in Phase 1. Two study team members conducted usability sessions in private rooms at community- and clinic-based study sites in four North Carolina cities. Each session lasted 60 to 90 minutes. Participants completed a pre-usability survey including demographics, disclosure experience (ever disclosed to a sexual partner; anal or vaginal sex without disclosure in past 12 months), use of technology, and VR experience (“Do you have any Virtual Reality experience? This can be through online virtual worlds, games, etc.”). Following use of the program, participants completed a semi-structured interview and a post-usability survey to rate the program on visual appeal, acceptability, ease of use, and user experience.

Each participant completed three practice disclosure dialogues. The first 15 participants used the manual system while the remaining 17 participants used both the manual system and the partially and fully animated versions of the program (Table 1). For each usability visit, the participant sat in one room using the program on a laptop while a study team member (interviewer) provided initial directions for use. The interviewer guided the participant through the program set-up choosing a virtual character partner and a disclosure setting (Figures 1.A and 1.B). The interviewer then prompted the participant with either Disclosure Scenario 1 (where the character has a neutral response to the participant) or Scenario 2 (where the character’s reaction is sympathetic) (Table 2). A second study team member (wizard) sat in a different room using an iPad tablet to either fully select the virtual character’s responses (manual version, Phase 2: Round 1) or ensure that the AI system’s selected response was appropriate (partially automated version, Phase 2: Rounds 2 and 3). When there was no appropriate utterance available within the database, the wizard typed a suitable response for the virtual character to deliver (Phase 2: Rounds 1 and 2).

After completing Scenarios 1 and 2, participants were given the option to complete a third dialogue (Scenario 3) with a virtual character that would respond negatively. The flow of the conversations within all three disclosure scenarios was similar. Conversations opened with small talk, led into disclosure and the post-disclosure reaction, and then concluded with addressing the future of the relationship or next steps. Due to the potential emotional impact of receiving a negative reaction when disclosing HIV status, study staff observed participants for any signs of distress during and after using the program. Only one disclosure dialogue was discontinued early due to participant reactivity. All practice disclosure dialogues were captured by the Tough Talks program as text-based output files and were used for further training the AI system, as described below.

Between each disclosure scenario dialogue and at the end of all three dialogues, participants were interviewed using a brief semi-structured guide to evaluate their experience using the program. In these interviews, participants were asked to assess the relevance and accuracy of the conversations as compared to their prior disclosure experiences (or anticipated experiences), emotional impact of using the program, how and where they could see themselves using the program, perceived usefulness of the program, and suggestions for improvements. Usability interview sessions were audio-recorded and professionally transcribed (verbalink.com).

Phase 2 Analysis—Usability session interview transcripts and disclosure dialogue transcripts were analyzed using the same process described in Phase 1. Analysis themes focused on the goal of refining the content of the disclosure scenarios, the virtual character’s reaction to participants’ disclosures, and the virtual character’s language and tone. Additional analysis themes focused on how the intervention was delivered including the reliability of the program producing intelligible disclosure dialogues and ease of program use.

Usability session notes, screen shots, and technical problems with the program were compiled during each usability session and sent to the technology partners for resolution. Bi-weekly analysis calls were held among the three partner organizations to discuss these notes and incorporate findings from usability testing into further refinement of the program. Microsoft Excel (Microsoft Office, 2016) was used to calculate descriptive statistics for participant survey data.

Phase 2 Results—Thirty-two HIV-positive YMSM (mean age 25, 78.1% Black/African American) completed 94 total practice disclosure scenario dialogues (each lasting between 5 and 15 minutes). Participants’ demographics and selected disclosure-related characteristics are presented in Table 3, Phase 2 column.

Most participants “agreed” or “strongly agreed” that Tough Talks was visually appealing, interesting and easy to use (Table 4). All participants “agreed” or “strongly agreed” that the program could help them have conversations around disclosing their HIV status.

Results from the usability sessions were incorporated into further refinement of the system to improve acceptability and functionality before the pilot trial (Phase 3). By the end of Phase 2, the study team had finalized the utterance database, sound recording and animation of all virtual character utterances, disclosure scenarios (Table 2, Column A), rules for the virtual characters’ reactions to disclosure (Table 2, Column B), and the full AI automation of the virtual characters’ responses for the neutral disclosure scenario.

Refinement of the natural language processing (NLP) system—As described in Technical Appendix A, the NLP system chooses the virtual character utterance in the fully automated system and makes suggestions for the semi-automated wizard-driven system. We used Phase 2 data to improve system performance to create the final version of the system that was used for the pilot study (Phase 3). The data from the last 26 usability test disclosure scenario dialogues in Phase 2 was withheld to use as a test set while the remaining data was used for more detailed linguistic analysis and improvement of the system.

Table 5, shows that across all three disclosure scenarios, the program’s ability to select “acceptable” utterances increased with each subsequent program dialogue manager. We annotated the training data using a coding scheme to indicate which virtual character utterances were direct responses to the participant versus the virtual character taking new initiative. We also identified utterances that referred to the overall context rather than the immediately prior utterance. For example, if a participant utterance was a simple answer, like “yes”, in order to ask a follow-up question, the virtual character would need to refer

back to its previous question, not just the participant's response to that question. Increased emphasis was placed on recognizing the disclosure event itself, since the new dialogue policy was very sensitive to this dialogue transition point. Incorporating these specifications led to a further increase in system performance, to an average of over 78% acceptable responses (range 71 — 84%, Table 5). Given resource constraints, a decision was made to focus subsequent program development for the automated prototype on the neutral disclosure dialogue scenario, which consistently had the highest AI performance, as shown in Table 5. This effort included tuning the dialogue policy to avoid repetition of system utterances, further improvement at recognizing disclosure events and taking dialogue initiative more often when the system did not understand a user utterance. In addition, the test data set of 26 dialogues was added to the final system used for the pilot (Phase 3).

Phase 3: Pilot study to assess program performance, acceptability, and preliminary efficacy of the Tough Talks prototype

Phase 3 Methods—The aim of Phase 3 was to conduct a technical pilot of the developed Tough Talks intervention prototype to assess the program's performance and its acceptability and preliminary efficacy at increasing HIV status disclosure self-efficacy among YMSM. Participants were recruited, screened for eligibility, and enrolled following the procedures described in Phases 1 and 2. Pilot-testing sessions were conducted using the same interview, survey and disclosure dialogue protocol as for Phase 2 with the following exceptions: 1) a fully automated version of the program was used for the neutral disclosure dialogue scenario such that the program operated without any wizard assistance; 2) the virtual character's speech function was activated such that the participant could both hear the virtual character's responses and see their typed responses on the screen; 3) participants completed the Self-efficacy to Disclose HIV Status scale (6 items, five-point response scale from "definitely could disclose" to "definitely could not disclose") (Parsons et al., 2005) pre- and post-intervention use and, 4) an adaptation of the Subjective Assessment of Speech System Interfaces (SASSI) scale was used to measure participant-rated acceptability and usability of the program (17 items, response categories on a five-point Likert scale of *strongly agree* to *strongly disagree*) (Hone & Graham, 2000). Each participant completed three or four program disclosure dialogues. As in Phase 2, disclosure dialogues were captured as output files by the Tough Talks' program. Pilot sessions lasted 60 to 90 minutes each and were also audio-recorded and transcribed following the methods described in Phase 2 above.

Phase 3 Analysis

Participant survey: Descriptive statistics were calculated for participant demographics and disclosure behaviors. Mean response scores were calculated for each item on the SASSI scale and the Self-efficacy to Disclose HIV Status scale.

System performance: All disclosure dialogue output files were reviewed to evaluate the virtual character's automated and suggested responses. For the sympathetic and negative scenarios, the percent of time in which the partially-automated AI system suggested the utterance chosen by the clinician wizard was calculated. For the neutral scenario, which used the fully automated AI system, utterances were evaluated by two members of the study staff on a scale from 0 (*inappropriate*) to 2 (*appropriate*), with the following meaning:

- Inappropriate (0): Response is problematic. The response does not make sense in context, such as asking for more, when the user was not saying anything, abruptly ending the dialogue, or asking the user a question that they have already answered;
- Not optimal but coherent (1): There is a better response utterance available. However, the response is not generally disruptive and allows the dialogue to continue, though perhaps not in the ideal direction.
- Appropriate (2): A quality response similar to what a wizard would have chosen and consistent with normal dialogue progression.

We also separately analyzed the performance of the system on correctly recognizing a disclosure event (e.g., “My HIV test was positive.”; “I was diagnosed with HIV two years ago.”) within each neutral disclosure dialogue. Discrepancies between evaluators were resolved by a third study team member.

Phase 3 Results

Demographics: The final automated program was tested with 11 new participants who completed a total of 32 program disclosure scenario dialogue sessions (mean age 25 years, 90.9% Black). Participant demographics and select characteristics are presented in Table 3, Phase 3 column.

Final system performance of semi-automated disclosure Scenarios 2 (sympathetic character response) and 3 (negative character response): Compared to the initial system from the Phase 2 usability, the intermediate AI system deployed in the Phase 3 pilot study improved its ability to suggest acceptable virtual character utterances for both the sympathetic and negative scenarios. Comparing Phase 2 to Phase 3, the number of program suggested utterances that the wizard actually selected rose from 17% to 34% in the sympathetic scenario, and from 17% to 26% in the negative scenario.

Final system performance of the fully automated disclosure Scenario 1 (neutral character response): There were 15 dialogues completed with the fully automated AI system in the neutral scenario, yielding a total of 286 computer-generated virtual character responses. Following the rating system described above, 203 (71.0%) were annotated as 2 (appropriate), 68 (23.8%) as 1 (not optimal but coherent), and 15 (5.2%) as 0 (inappropriate). Compared to prior versions of the program, the final AI system for the neutral disclosure scenario showed significant improvements in selecting appropriate disclosure utterances (two-sample: $t = 3.557$, $df = 570$, $p < 0.001$).

In all 15 neutral dialogues using the final AI system, the system was able to detect an HIV status disclosure by the participant. Only 3 of 15 dialogues (five utterances overall) included a disclosure-related inappropriate utterance.

Participant evaluation of Tough Talks: On average, the program received favorable or neutral responses for most categories on the SASSI scale (Table 6). Most (9/11) participants agreed or strongly agreed that the system was easy to use, that they would like to use the system

frequently (9/11), and that most people would learn to use this system very quickly (11/11). Around half of participants thought the system was unpredictable (5/11) or did not always do what was expected (6/11).

We did not see any statistically significant changes in disclosure self-efficacy during the pilot trial (Table 7).

Discussion

While disclosure offers the possibility of individual and public health benefits for HIV care and prevention (O'Connell et al., 2015; Stirratt et al., 2006), YMSM experience HIV status disclosure as a complex personal and social process with implications for fear of exposure and persistent HIV stigma within and outside of the gay community (J. Bird et al., 2017; J. D. Bird & Voisin, 2013). Our work with YMSM in developing the Tough Talks program confirmed these known disclosure barriers but also identified expressed need and enthusiasm for a program that could support youth in thinking through disclosure decisions and developing communication skills and disclosure self-efficacy. This approach to supporting HIV status disclosure acknowledges the fraught nature of disclosure decisions while also adopting an empowerment lens that prioritizes the agency of individual YMSM. A similar strategy could be applied to other topics that impact health and wellbeing - from discussing one's sexual or gender identity, to training doctors to improve delivery of difficult diagnoses and help patients talk with their families about a variety of medical conditions.

Through the development of the Tough Talks VR prototype, we were able to demonstrate proof of concept that AI driven scenarios are both feasible and acceptable for YMSM to practice HIV status disclosure with sex partners. It is not surprising that we did not see changes in disclosure self-efficacy given the limited ability for participants to practice within the VR scenarios as well as our small sample size. Future developments within the program will continue to build on the theoretical foundations of SCT to include additional knowledge content, behavioral modeling, skills-building exercises to increase disclosure self-efficacy, and self-reflection activities along with the practice disclosure dialogues. We will also include skills-building modules on critical assessments of disclosure scenarios, as not all disclosure contexts may be safe or desirable for direct disclosure.

The development of the Tough Talks program highlights the importance of end-user engagement in eHealth intervention development (Dumas & Redish, 1999; Noar, Benac, & Harris, 2007; usability.gov, 2017). We involved YMSM and solicited feedback at every phase in development through focus groups, one-on-one usability, and pilot-testing sessions. YMSM informed the content and features of the program, reflected on their own disclosure experiences, and provided insight into how they experienced the intervention. We iteratively analyzed their feedback, made changes to the program, and continued testing it to reach acceptability and feasibility. This process was particularly important for developing functional, realistic VR, AI and NLP components as the program needed to simulate a disclosure experience that was believable and engaging for participants as well as be able to recognize and appropriately respond to participants' natural styles of communication. Based on participant feedback, the next version of Tough Talks will include expanded program

ability via more virtual character options, disclosure settings, and disclosure scenarios. We will also further develop the coach feature of the program to give it the capacity to guide the user through a disclosure experience.

Disclosure dialogues can be quite nuanced; creating a virtual character that understood what participants typed and responded appropriately within the complexity of these dialogues was challenging. There were three main challenges for the Tough Talks AI system. First, there was a small amount of training data available. AI systems built with the NPCEditor (the system used to create realistic automated dialogue interactions) require adequate training data from the target population to compute effective translation and language models (Leuski & Traum, 2008, 2010, 2011). Results from Phase 2 testing show that the amount of data we collected (94 dialogues from 32 participants, containing about one thousand participant utterances) was not sufficient to achieve adequate system performance across all dialogue phases and disclosure scenarios. Second, the mixed-initiative nature of disclosure dialogues, including both virtual character reactions to the participant and virtual character initiations of conversation topics, is different from the typical participant-driven scenarios at which the NPCEditor classifier excels (Leuski & Traum, 2010). Specifically, in Tough Talks disclosure dialogues, it is often necessary for the virtual character to move beyond simple responses to a participant and take initiative in the conversation using questions and unprompted statements. Finally, complex conversation policies were needed for different phases of disclosure dialogues and different disclosure scenarios. While the NLP component steadily improved through the usability testing phase, further data collection, analysis, and system development is needed to create a fully automated AI system that can reliably cope with a wider range of participants' experiences and conversational styles, particularly for the sympathetic and negative scenarios. This finding is relevant for the broader field in terms of considering the function and feasibility of using automated AI systems for other HIV disclosure scenarios (e.g., to family members, friends, or providers) as well as other health or education settings. Dialogue topics and tasks that are more complex, have a larger number of possible outcomes, and require more initiative on the part of the AI will require significantly larger amounts of training data to ensure robust, productive conversations.

The wizard-driven version of the Tough Talks program was intended as a stepping stone in the iterative development cycle toward a fully automated, interactive system. However, given the complexity of some disclosure dialogues and how well-received the program was by participants, this version of the program could be meritorious as an intervention on its own. While it would not have the one-to-many power of a fully automated system since it would need a human operator, the VR experience would still offer advantages over didactic instruction and standard in-person role play. Namely, the sensory-rich immersive environments of VR (virtual characters, visual ambience, directional audio, culturally specific content) provide a realistic alternative for YMSM for *in vivo* rehearsal and performance of HIV disclosure behaviors in a controlled environment where new challenges can be gradually introduced. Further, VR environments provide a standardized setting that can be controlled and replicated to deliver an intervention in a systematic manner. As emotional release/catharsis have been identified as one core reason for disclosure (Serovich & Mosack, 2003), it is also worth exploring whether the program could offer a therapeutic benefit to the end user regardless of whether or not they go on to disclose in the real world.

This effect could serve as an end in itself or function as a stepping stone toward future consideration of disclosure - a feature that may be particularly relevant for YMSM who are not yet or not currently sexually active (Thoth et al., 2014).

Acknowledgements

We thank all study participants who shared their time and experiences to shape the development of this program. Thank you to Em C Pike and Meera Ravi for providing research assistance in the early phases of program development. We acknowledge the numerous artists and developers at Virtually Better Inc. and the University of Southern California Institute for Creative Technologies who contributed their talent and expertise to this prototype. The research reported in this publication was supported by the National Institute of Mental Health under Award Number R43MH104102. The content is solely the responsibility of the authors and the views expressed herein do not reflect the official stance of any funding agency. We have no conflicts of interests to declare.

Technical Appendix A

Virtual Character Development

Developing the virtual characters for the Tough Talks project was an involved and intricate process that used multiple specialized programs. Once character designs were approved after feedback from the focus groups, an artist built the base character model in Autodesk Maya (Autodesk, Inc., San Rafael, CA) standing in a neutral pose. The character model was then loaded into Autodesk Mudbox (Autodesk, Inc., San Rafael, CA), where artists sculpted in finer details such as creases and wrinkles on the face or clothing, skin pores, as well as painted the character's clothing and skin textures. Adding these details in Mudbox adds millions of polygons, far beyond what a game engine can handle in real-time, but the fine details can be converted into texture data. The final result of this process is that the low-resolution, game ready model (12,000 polygons) is visually indistinguishable from what was developed in Mudbox (>2,000,000 polygons). The character model was then put back into Maya for rigging, where the artist built the functional skeleton for the character. The process was similar to stringing up a puppet: points of articulation were defined and a determination made for the way each joint could move. This included building a control rig for the face so expressions and lipsync could be animated.

With the character assets ready for animation, the dialog utterances from the approved database were recorded. The audio was then cleaned (to remove ambient noise and random clicks and pops in the voice actor's speech), and the individual utterances separated for animation. Animating the lips for speech is typically a very slow and arduous process, involving the use of a middleware program called FaceFX (OC3 Entertainment, Ocean Grove, NJ) which analyzes the audio file and animates the lips dynamically. FaceFX has been widely employed in the gaming industry due to its high-quality results and reliable integration with animation software and game engines; for Tough Talks this saved weeks of development time. While lip movement is automated, body animation is done manually for each line of dialog so that the visual performance matches the vocal utterances. Once the animation for 126 individual lines of dialog (virtual character utterance database) was completed, the virtual character was exported to Unity (Unity Technologies, San Francisco, CA) to be placed in the environment. As the participant player and camera are stationary, to load in the full 3D environments would have been inefficient. Instead, still images of the

scenes (bar and park) were utilized through placement behind the virtual character, much like a billboard.

Natural Language Processing/AI System Development

The main task of the AI system was to choose the virtual character utterances for the fully automated system. A subordinate function was added to make suggestions for a semiautomated system, in which a clinician wizard could choose to use these suggestions or chose something else. The goal of the system was not to fully replicate the performance of clinician wizards, but to make the dialogues realistic enough to allow the participants to have a constructive experience in practicing disclosure, without requiring the active participation of a clinician in the interaction.

The system used to implement this functionality was the NPCEditor (Leuski & Traum, 2011). This system is part of the virtual human toolkit (Hartholt et al., 2013), and has been previously used to create several very successful dialogue interactions (Artstein et al., 2008; Robinson, Traum, Ittycheriah, & Henderer, 2008; Traum et al., 2012; Traum et al., 2015). The NPCEditor uses a cross-language relevance model to choose a response utterance, given an input utterance, and a set of training data from that domain that indicates some good inputs and responses. The approach creates a translation model from input words to a language model for outputs, which can be compared to the available outputs (Leuski & Traum, 2008). Another component is a dialogue manager, that uses programmed rules that can operate on recognized new input features as well as stored prior dialogue context to decide how to deploy the classifier (e.g. which domain or data set to apply when, and what to do if there is no utterance above the threshold, or if the best answer has already been said recently).

Several adaptations were made for the NPCEditor AI system to communicate with the Tough Talks front-end program, including triggering virtual character behavior. A web-based communication protocol was set up, such that clinicians in several locations could easily access the most up-to-date version of the system. The system also logged all inputs and outputs for later analysis.

Biography

Kathryn E. Muessig is an Assistant Professor in the Department of Health Behavior at the University of North Carolina at Chapel Hill's (UNC) Gillings School of Global Public Health. Dr. Muessig uses behavioral science, gamification theory, and an anthropological lens to develop and evaluate eHealth and mHealth interventions for HIV prevention and care. She leads a mixed methods study of an online intervention for stigma reduction among young Black men who have sex with men and pilot studies in China and Vietnam to reduce stigma and improve HIV continuum of care outcomes.

Contact: kmuessig@med.unc.edu

Kelly A. Knudtson is a Project Manager at the Institute of Global Health and Infectious Disease at UNC. She received her Master's in Public Health in Health Behavior at UNC.

Her current research interests focus on implementing harm reduction interventions, improving LGBTQ health and well-being, addressing health disparities, and using technology as a tool to empower individuals to advocate for their health. She has expertise in qualitative research methods and analysis, working directly with vulnerable populations, and geographic information systems (GIS).

Karina Soni was a Project Coordinator at the UNC Behavior and Technology Lab from 2014 — 2018. In this role she supported formative and evaluative research on a number of web, mobile web and mobile application-based interventions to improve HIV prevention and care. She received a BA from UNC Chapel Hill with a double major in journalism and public policy and is now pursuing a Master of Science in Information Science at UNC Chapel Hill. She has expertise in graphic design, visual display of information, qualitative research methods, and conducting research with community stakeholders.

Margo Adams Larsen is Director of Research at Virtually Better, Inc., a technology company in Decatur, Georgia, dedicated to advancing the use of technology in behavioral healthcare. Dr. Adams Larsen has special interests in the use of technology in clinical practice, pediatric psychology, and empirically validated interventions. Her current research interests focus on technology-based interventions targeting HIV risk reduction and pediatric anxiety disorders.

David Traum is a Research Assistant Professor and principal scientist at the University of Southern California (USC) Institute for Creative Technologies (ICT), and a scientific advisory board member and past president of SIGDIAL (international special interest group in discourse and dialogue). At ICT, Dr. Traum leads the Natural Language Dialogue Group, which researches all aspects of natural language (NL) dialogue, including dialogue management, spoken and NL understanding and generation and dialogue evaluation. His group collaborates on projects involving integrated virtual humans and transitioning NL dialogue capability into interactive training applications. Dr. Traum has pioneered research in computational dialogue modeling, including how common ground is established through conversation, the information state approach to dialogue, multiparty dialogue, and noncooperative dialogue. He is a founding editor of the Journal of Dialogue and Discourse.

Willa Dong is a PhD student in the Department of Health Behavior at UNC's Gillings School of Global Public Health. She received her Master's of Science in Public Health from the Johns Hopkins Bloomberg School of Public Health. Her research interests include stigma, global LGBT health, and community-based participatory research methods. She has conducted research in multiple settings in China and the U.S. and has expertise in the use of qualitative and ethnographic data collection and analysis methods and working with multiple stakeholder groups (e.g. governments, universities, and community-based organizations) to conduct research studies.

Donaldson F. Conserve is an Assistant Professor in the Arnold School of Public Health at the University of South Carolina. His research focuses on developing, implementing, and evaluating novel intervention strategies to prevent and treat HIV among men. He has particular interest in the areas of novel strategies for promoting HIV testing (e.g. self-testing)

and understanding the complexities of HIV status disclosure. Dr. Conserve has conducted public health research in the U.S. and global settings including Tanzania, Haiti, South Africa, Nigeria and Europe.

Anton Leuski is a Research Assistant Professor at the USC Department of Computer Science and a research scientist at USC ICT. He received his Ph.D. in computer science from the University of Massachusetts at Amherst. His research interests include interactive information access, human-computer interaction, and machine learning. Dr. Leuski's recent work has focused on applying statistical information retrieval approaches to dialogue text analysis, natural language understanding and generation. Of primary interest is the development of the statistical approaches for building effective text classification techniques with small amounts of training data. These text classification approaches facilitate rapid development of practical solutions for natural language dialogue applications.

Ron Artstein is a Research Assistant Professor with USC ICT and the USC School of Engineering Computer Science Department. Dr. Artstein is a linguist and computational linguist whose training and early work was in theoretical linguistics, specializing in formal semantics. His primary research interests are in corpus creation and management, reliability statistics for corpus annotation, dialogue, anaphora resolution, and the formal semantics of natural language. He also works on coordinating an effort to create and organize a corpus of spoken dialogue which is used in developing language comprehension and speech of virtual humans.

Lisa B. Hightow-Weidman is an Associate Professor of Medicine and Health Behavior at UNC with more than 15 years of experience working with technology in HIV-infected and at-risk populations. She is the founding Director of UNC's Behavior and Technology Lab which is home to a large portfolio of externally-funded research initiatives. Her research focuses on technology-based interventions to prevent HIV infection and improve linkage and retention in care for populations at high risk of acquiring or transmitting HIV, particularly young men who have sex with men. She is an active clinical HIV provider in a large HIV-clinic managing at-risk and HIV-infected youth and young adults.

Contact: lisa.hightow@med.unc.edu

References

- Abler L , Sikkema KJ , Watt MH , Hansen NB , Wilson PA , & Kochman A (2015). Depression and HIV Serostatus Disclosure to Sexual Partners Among Newly HIV-Diagnosed Men Who Have Sex with Men. *AIDS Patient Care STDS*, 29(10), 550–558. doi:10.1089/apc.2015.0122 [PubMed: 26430721]
- Ahn SJ (2015). Incorporating Immersive Virtual Environments in Health Promotion Campaigns: A Construal Level Theory Approach. *Health Commun*, 30(6), 545–556. doi: 10.1080/10410236.2013.869650 [PubMed: 24991725]
- Artstein R , Gandhe S , Leuski A , & Traum D (2008, May 2008). Field Testing of an Interactive Question-Answering Character. Paper presented at the Language Resources and Evaluation Conference (LREC), Marrakech, Morocco.
- Bachmann LH , Grimley DM , Gao H , Aban I , Chen H , Raper JL , . . . Hook EW (2013). Impact of a computer-assisted, provider-delivered intervention on sexual risk behaviors in HIV-positive men

- who have sex with men (MSM) in a primary care setting. *AIDS Educ Prev*, 25(2), 87–101. doi: 10.1521/aeap.2013.25.2.87 [PubMed: 23514077]
- Bandura A (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Bandura A (2011). Social cognitive theory In Van Lange P , Kruglanski A , & Higgins T (Eds.), *The Handbook of Theories of Social Psychology* (pp. 349–373). Thousand Oaks, CA: Sage.
- Bird J , Eversman M , & Voisin D (2017). “You just can’t trust everybody”: the impact of sexual risk, partner type and perceived partner trustworthiness on HIV-status disclosure decisions among HIV-positive black gay and bisexual men. *Cult Health Sex*, 19(8), 829–843. doi: 10.1080/13691058.2016.1267408 [PubMed: 28050947]
- Bird JD , & Voisin DR (2013). “You’re an open target to be abused”: a qualitative study of stigma and HIV self-disclosure among Black men who have sex with men. *Am J Public Health*, 103(12), 2193–2199. doi:10.2105/ajph.2013.301437 [PubMed: 24134345]
- Broadus MR , DiFranceisco WJ , Kelly JA , St Lawrence JS , Amirkhanian YA , & Dickson-Gomez JD (2015). Social Media Use and High-Risk Sexual Behavior Among Black Men Who Have Sex with Men: A Three-City Study. *AIDS Behav*, 19 Suppl 2, 90–97. doi:10.1007/s10461-014-0980-z [PubMed: 25566762]
- Chaudoir SR , Fisher JD , & Simoni JM (2011). Understanding HIV disclosure: a review and application of the Disclosure Processes Model. *Soc Sci Med*, 72(10), 1618–1629. doi:10.1016/j.socscimed.2011.03.028 [PubMed: 21514708]
- Chiasson MA , Shaw FS , Humberstone M , Hirshfield S , & Hartel D (2009). Increased HIV disclosure three months after an online video intervention for men who have sex with men (MSM). *AIDS Care*, 21(9), 1081–1089. doi:10.1080/09540120902730013 [PubMed: 20024766]
- Christensen JL , Miller LC , Appleby PR , Corsbie-Massay C , Godoy CG , Marsella SC , & Read SJ (2013). Reducing shame in a game that predicts HIV risk reduction for young adult MSM: a randomized trial delivered nationally over the Web. *J Int AIDS Soc*, 16(3 Suppl 2), 18716. doi: 10.7448/ias.16.3.18716 [PubMed: 24242264]
- Conserve DF , Groves AK , & Maman S (2015). Effectiveness of interventions promoting HIV serostatus disclosure to sexual partners: a systematic review. *AIDS Behav*, 19(10), 1763–1772. doi: 10.1007/s10461-015-1006-1 [PubMed: 25645328]
- Cook SH , Valera P , & Wilson PA (2015). HIV status disclosure, depressive symptoms, and sexual risk behavior among HIV-positive young men who have sex with men. *J Behav Med*, 38(3), 507–517. doi:10.1007/s10865-015-9624-7 [PubMed: 25773478]
- Dahlback N , Jonsson A , & Ahrenberg L (1993). Wizard of Oz studies - Why and How. *Knowledge-Based Systems*, 6(4), 258–266. doi:10.1016/0950-7051(93)90017-N
- Dima AL , Stutterheim SE , Lyimo R , & de Bruin M (2014). Advancing methodology in the study of HIV status disclosure: the importance of considering disclosure target and intent. *Soc Sci Med*, 108, 166–174. doi:10.1016/j.socscimed.2014.02.045 [PubMed: 24641881]
- Dumas JS , & Redish JC (1999). *A Practical Guide to Usability Testing*. Portland, Oregon: Intellect Books.
- Durlach N , & Mavor A (Eds.). (1995). *Virtual Reality: Scientific and Technological Challenges*. Washington DC: National Academy Press.
- Grov C , Agyemang L , Ventuneac A , & Breslow AS (2013). Navigating condom use and HIV status disclosure with partners met online: a qualitative pilot study with gay and bisexual men from Craigslist.org. *AIDS Educ Prev*, 25(1), 72–85. doi:10.1521/aeap.2013.25.1.72 [PubMed: 23387953]
- Grov C , Rendina HJ , Moody RL , Ventuneac A , & Parsons JT (2015). HIV Serosorting, Status Disclosure, and Strategic Positioning Among Highly Sexually Active Gay and Bisexual Men. *AIDS Patient Care STDS*, 29(10), 559–568. doi:10.1089/apc.2015.0126 [PubMed: 26348322]
- Hartholt A , Traum D , Marsella S , Shapiro A , Giota S , Leuski A , . . . Gratch J (2013, August 29–31). All Together Now, Introducing the Virtual Human Toolkit. Paper presented at the International Conference on Intelligent Virtual Agents, Edinburgh, UK.
- Hightow-Weidman LB , Phillips G , Outlaw AY , Wohl AR , Fields S , Hildalgo J , & LeGrand S (2013). Patterns of HIV disclosure and condom use among HIV- infected young racial/ethnic

- minority men who have sex with men. *AIDS Behav*, 17(1), 360–368. doi:10.1007/s10461-012-0331-x [PubMed: 23054043]
- Hone K , & Graham R (2000). Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). *Nat. Tang. Eng*, 6(3–4), 287–303. doi:10.1017/s1351324900002497
- Huang H-M , Rauch U , & Liaw S-S (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education*, 55(3), 1171–1182. doi:10.1016/j.compedu.2010.05.014
- Kalichman SC , Kalichman MO , Cherry C , & Grebler T (2016). HIV Disclosure and Transmission Risks to Sex Partners Among HIV Positive Men. *AIDS Patient Care STDS*, 30(5), 221–228. [PubMed: 27158850]
- Leuski A , & Traum D (2008, December). A Statistical Approach for Text Processing in Virtual Humans. Paper presented at the 26th Army Science Conference.
- Leuski A , & Traum D (2010). NPCEditor. A Tool for Building Question-Answering Characters Paper presented at the Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, Valletta, Malta.
- Leuski A , & Traum D (2011). NPCEditor: Creating virtual human dialogue using information retrieval techniques. *AI Magazine*, 32(2), 42–56.
- Li H , Chen X , & Yu B (2016). Disclosure appraisal mediating the association between perceived stigma and HIV disclosure to casual sex partners among HIV+ MSM: a path model analysis. *AIDS Can*, 28(6), 722–725. doi:10.1080/09540121.2016.1140884
- Mausbach BT , Semple SJ , Strathdee SA , Zians J , & Patterson TL (2007). Efficacy of a behavioral intervention for increasing safer sex behaviors in HIV-positive MSM methamphetamine users: results from the EDGE study. *Drug Alcohol Depend*, 87(2–3), 249– 257. doi:10.1016/j.drugalcdep.2006.08.026 [PubMed: 17182196]
- Milam J , Morris S , Jain S , Sun X , Dube MP , Daar ES , . . . Haubrich R (2016). Randomized Controlled Trial of an Internet Application to Reduce HIV Transmission Behavior Among HIV Infected Men Who have Sex with Men. *AIDS Behav*, 20(6), 1173– 1181. doi:10.1007/s10461-015-1215-7 [PubMed: 26487300]
- Morie JR , Antonisse J , Bouchard S , & Chance E (2009). Virtual worlds as a healing modality for returning soldiers and veterans. *Stud Health Technol Inform*, 144, 273–276. [PubMed: 19592780]
- Morie JR , & Chance E (2011). Extending the reach of health care for obesity and diabetes using virtual worlds. *J Diabetes Set Technol*, 5(2), 272–276.
- Morin SR , Steward WT , Charlebois ED , Remien RH , Pinkerton SD , Johnson MO , . . . Chesney MA (2005). Predicting HIV transmission risk among HIV-infected men who have sex with men: findings from the healthy living project. *J Acquir Immune Defic Syndr*, 40(2), 226–235. [PubMed: 16186742]
- Muessig KE , Baltierra NB , Pike EC , LeGrand S , & Hightow-Weidman LB (2014). Achieving HIV risk reduction through HealthMpowerment.org, a user-driven eHealth intervention for young Black men who have sex with men and transgender women who have sex with men. *Digit Cult Educ*, 6(3), 164–182. [PubMed: 25593616]
- Muessig KE , LeGrand S , Sanon N , Soni K , Nwoko N , Platt A , & Hightow-Weidman LB (2017). 'Yes, I'm ready, and I'm taking them!': HIV-infected young men who have sex with men and transgender women evaluate the Epic Allies gamified smartphone app for improving antiretroviral uptake and adherence. Paper presented at the International Association of Providers of AIDS Care, 12th International Conference on HIV Treatment and Prevention Adherence (Adherence 2017), Miami, Florida, U.S.A.
- Muessig KE , Pike EC , Fowler B , LeGrand S , Parsons JT , Bull SS , . . . Hightow-Weidman LB (2013). Putting prevention in their pockets: developing mobile phone-based HIV interventions for black men who have sex with men. *AIDS Patient Care STDS*, 27(4), 211–222. doi:10.1089/apc.2012.0404
- Noar SM , Benac CN , & Harris MS (2007). Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. *Psychol Bull*, 133(4), 673–693. doi: 10.1037/0033-2909.133.4.673 [PubMed: 17592961]

- O'Connell AA , Reed SJ , & Serovich JA (2015). The efficacy of serostatus disclosure for HIV Transmission risk reduction. *AIDS Behav*, 19(2), 283–290. doi:10.1007/s10461-014-0848-2 [PubMed: 25164375]
- Overstreet NM , Eamshaw VA , Kalichman SC , & Quinn DM (2013). Internalized stigma and HIV status disclosure among HIV-positive black men who have sex with men. *AIDS Care*, 25(4), 466–471. doi:10.1080/09540121.2012.720362 [PubMed: 23006008]
- Parsons JT , Schrimshaw EW , Bimbi DS , Wolitski RJ , Gomez CA , & Halkitis PN (2005). Consistent, inconsistent, and non-disclosure to casual sexual partners among HIV-seropositive gay and bisexual men. *Aids*, 19 Suppl 1, S87–97. [PubMed: 15838198]
- Persson A , & Richards W (2008). From closet to heterotopia: a conceptual exploration of disclosure and 'passing' among heterosexuals living with HIV. *Cult Health Sex*, 10(1), 73– 86. doi: 10.1080/13691050701669048 [PubMed: 18038282]
- Pinkerton SD , Holtgrave DR , & Galletly CL (2008). Infections prevented by increasing HIV serostatus awareness in the United States, 2001 to 2004. *J Acquir Immune Defic Syndr*, 47(3), 354–357. doi:10.1097/QAI.0b013e318160d57e [PubMed: 18176322]
- Przybyla SM , Golin CE , Widman L , Grodensky CA , Earp JA , & Suchindran C (2013). Serostatus disclosure to sexual partners among people living with HIV: examining the roles of partner characteristics and stigma. *AIDS Care*, 25(5), 566–572. doi:10.1080/09540121.2012.722601 [PubMed: 23020136]
- Riva G , Mantovani F , Capideville CS , Preziosa A , Morganti F , Villani D , . . . Alcaniz M (2007). Affective interactions using virtual reality: the link between presence and emotions. *Cyberpsychol Behav*, 10(1), 45–56. doi:10.1089/cpb.2006.9993 [PubMed: 17305448]
- Robinson S , Traum D , Ittycheriah M , & Henderer J (2008, May 26 - June 1). What would you ask a Conversational Agent? Observations of Human-Agent Dialogues in a Museum Setting. Paper presented at the Language Resources and Evaluation Conference (LREC), Marrakech, Morocco.
- Ronn M , White PJ , Hughes G , & Ward H (2014). Developing a conceptual framework of seroadaptive behaviors in HIV-diagnosed men who have sex with men. *J Infect Dis*, 210 Suppl 2, S586–593. doi:10.1093/infdis/jiu482 [PubMed: 25381379]
- Serovich JM , Laschober TC , Brown MJ , & Kimberly JA (2017). Evaluation of HIV Disclosure Behavior Following a Randomized Controlled Disclosure Intervention for Men Who Have Sex with Men Living with HIV. *Arch SexBehav*. doi:10.1007/s10508-017-1055-9
- Serovich JM , & Mosack KE (2003). Reasons for HIV disclosure or nondisclosure to casual sexual partners. *AIDS Educ Prev*, 15(1), 70–80. [PubMed: 12627744]
- Serovich JM , Reed S , Graftsky EL , & Andrist D (2009). An intervention to assist men who have sex with men disclose their serostatus to casual sex partners: results from a pilot study. *AIDS EducPrev*, 21(3), 207–219. doi:10.1521/aeap.2009.21.3.207
- Stirratt MJ , Remien RH , Smith A , Copeland OQ , Dolezal C , & Krieger D (2006). The role of HIV serostatus disclosure in antiretroviral medication adherence. *AIDS Behav*, 10(5), 483–493. doi: 10.1007/s10461-006-9106-6 [PubMed: 16721505]
- Thoth CA , Tucker C , Leahy M , & Stewart SM (2014). Self-disclosure of serostatus by youth who are HIV-positive: a review. *J Behav Med*, 37(2), 276–288. doi:10.1007/s10865-012-9485-2 [PubMed: 23277232]
- Traum D , Aggarwal P , Artstein R , Foutz S , Gerten J , Katsamanis A , . . . Swartout W (2012). Ada and Grace: Direct interaction with museum visitors. Paper presented at the Intelligent Virtual Agents (IVA-2012), Santa Cruz, California, USA.
- Traum D , Jones A , Hays K , Maio H , Alexander O , Artstein R , ... Swartout W (2015). New Dimensions in Testimony: Digitally Preserving a Holocaust Survivor's Interactive Storytelling. Paper presented at the 8th International Conference on Interactive Digital Storytelling (ICIDS), Copenhagen, Denmark.
- U. S. Centers for Disease Control and Prevention. (2014). Start Talking. Stop HIV. Preventing New HIV Infections through Open Conversations. Retrieved from <https://www.cdc.gov/nchhstp/newsroom/docs/factsheets/stsh-factsheet-508.pdf>

U.S. Centers for Disease Control and Prevention. (2016). HIV Surveillance Report, 2015. 27. Retrieved from <https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2015-vol-27.pdf>

usability.gov. (2017). User-Centered Design Basics. Retrieved from <https://www.usability.gov/what-and-why/user-centered-design.html>

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

a. Virtual characters



b. Virtual disclosure settings



c. Tough Talks Wizard Interface iPad Screenshot

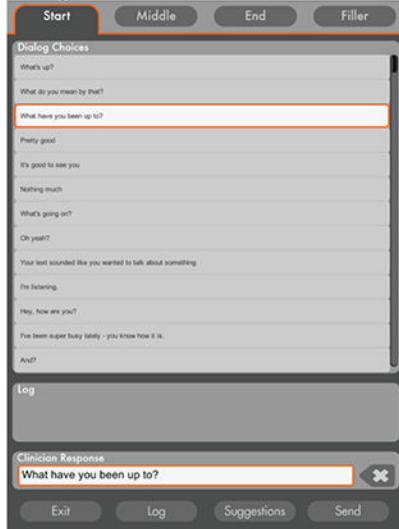


Figure 1. Screenshots of Tough Talks Virtual Reality Program

Figure 1 images depict program screenshots: virtual characters chosen as potential disclosure partners (a), settings chosen to practice disclosing (b), and the wizard interface in which new lines of dialogue could be written (c).

Table 1. Tough Talks HIV status disclosure virtual reality program development and research methods

Study Phase:	Phase 2: Iterative Development and Usability Testing			Phase 3: Pilot Trial
	Round 1	Round 2	Round 3	
Phase 1: Formative Research	<p>Focus groups that included disclosure role plays and assessment of visual program assets</p> <p>4 focus groups (7 HIV-negative MSM; 8 HIV-positive MSM)</p> <p>6 in-person disclosure role play dialogues</p>	<p>Usability sessions with initial program, semi-structured user-experience interview, prepost surveys</p> <p>15 HIV-positive MSM</p>	<p>15 HIV-positive MSM</p>	<p>Pilot testing with refined program, semi-structured user experience interviews, prepost surveys</p> <p>11 HIV-positive MSM</p>
Data Collection Methods:				
Total Participants: (n=58)				
Total Disclosure dialogues: (n=132)	44 program disclosure dialogues	42 program disclosure dialogues	8 program disclosure dialogues	32 program disclosure dialogues
Program visuals: virtual character, disclosure settings, virtual disclosure coach	<ul style="list-style-type: none"> - Virtual character build (nonspeaking, immobile, neutral pose, refined details and features of face, skin, clothing) - Rig build (controls for facial expressions and lipsync) 	<ul style="list-style-type: none"> - Audio utterance recordings - Speaking virtual character (lip automation) - Expressive virtual character (character face and body animation) 	<ul style="list-style-type: none"> - Coach feature programmed (speech bubbles with audio and verbal cues for end user) 	<ul style="list-style-type: none"> - Refined all built features for smooth integration with AI and NLP engines
Natural language processing: artificial intelligence	<ul style="list-style-type: none"> - All virtual character responses manually selected or typed by clinician wizard - Refined virtual character utterance database (n=114 utterances) 	<ul style="list-style-type: none"> - AI program suggests virtual character responses, clinician wizard agrees or overrides with manual selection 	<ul style="list-style-type: none"> - Neutral disclosure scenario: Virtual character responses fully automated using AI - Positive/negative disclosure scenarios: Program suggests virtual character responses, wizard agrees or overrides - Finalized virtual character utterance database (n=125 utterances) 	<ul style="list-style-type: none"> - Neutral disclosure scenario: Virtual character responses fully automated using AI - Positive/negative disclosure scenarios: Program suggests virtual character responses, wizard agrees or overrides - Employed finalized virtual character utterance database (n=125 utterances)

Table 2.

Tough Talks program disclosure scenarios with virtual character reaction

	A: Disclosure scenario description	B: Virtual character reaction	C: Sample virtual character utterance
Scenario 1	Participant has been seeing the virtual character for six weeks. They have not had sex yet, but the relationship seems to be going there and the participant wants to disclose their HIV-positive status.	Neutral: responses are neither strongly positive nor strongly negative. Character expresses curiosity and uncertainty.	“I need time to think about everything.” “Do I need to get tested?”
Scenario 2	Participant has been dating the virtual character seriously for six months and they have been having condomless sex. The participant gets tested and finds out they are HIV positive and wants to tell their partner.	Sympathetic: responses are generally positive. Character expresses some concern for his own potential risk, but is ultimately sympathetic, wants to provide support, and is invested in the relationship.	“I’m just grateful you’re telling me now.” “We can get through this together.”
Scenario 3	Participant has met the virtual character through an online hookup app and they have been having condomless sex for a few weeks. The participant finds out they are HIV positive and decides to tell this partner.	Negative: responses are negative and confrontational. Character expresses disinterest in relationship and anger for being put at risk.	“I can’t believe you’re telling me this.” “Why didn’t you tell me this before!?”

R (finement of the natural language processing (NLP) system)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Participant demographics and select characteristics, Phases 2 and 3

	Phase 2 Usability testing (n=32)	Phase 3 Pilot testing (n=11)
Age mean (range)	25 (18–29)	25 (20 – 28)
Race/ethnicity		
Asian	1	0
Black/African American	25	10
Hispanic/Latino	1	1
Native American	1	0
White	4	0
Sexual orientation		
Gay	25	6
Bisexual	7	5
Education		
Some high school	2	1
Completed high school/GED	10	2
Some college	14	3
Some Professional, Technical or Trade school	2	0
College degree or more	4	5
Employment		
Employed	23	3
Unemployed	9	8
Current relationship status		
Committed relationship with a man/men	15	4
Other relationship with a man/men	1	1
Not in a relationship	16	6
Any anal/vaginal sex, past 12 months with a partner that had not disclosed to?	7	7
Never disclosed HIV status to a sexual partner, before or after sex	4	1
Any prior experience using virtual reality programs	19	7

Table 4.

Acceptability of the Tough Talks program among 32 HIV-infected MSM, Phase 2

Survey item	Mean score
The program was interesting.	1.13
The program was visually appealing.	1.73
The program functions were easy to use.	1.00
My overall impression of the program was favorable.	1.33
The language in the program reflected language I might use or hear.	1.67
I felt comfortable interacting with the virtual character(s) in this program.	1.73
I found the virtual character I was talking to visually appealing.	1.83
The program layout and structure was easy to understand and navigate.	1.53
I felt comfortable talking about HIV in this setting. (The park)	1.53
I felt comfortable talking about HIV in this setting. (The bar)	3.00
This program could help me have conversations around HIV with another person.	1.33
I can see myself using something like this when it's completed.	1.33

Averaged score on 5 point Likert scale reported, (1=strongly agree; 2=agree; 3=neither agree nor disagree; 4=disagree; 5=strongly disagree)

Table 5.

Iterative improvement of the Tough Talks AI system as applied to a test set of 26 participant dialogues from Phase 2 usability sessions

Row		Disclosure Scenario		
		Neutral	Negative	Positive
1	Average percentage of AI suggested virtual character utterances selected by wizard using original dialogue manager	25%	17%	17%
2	Average percentage of AI suggested virtual character utterances coded as “acceptable” using refined dialogue manager including dialogue policy	54%	51%	49%
3	Average percentage of AI suggested virtual character utterances coded as “acceptable” using further refined dialogue manager (re-annotation and disclosure analysis)	84%	80%	71%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6.

Subjective Assessment of Speech System Interfaces (SASSI) for Tough Talks program among 11 HIV-infected MSM, age 18–30

Survey item	Mean score
I think that I would like to use this system frequently.	2.27
I found the system unnecessarily complex.	4.18
I thought the system was easy to use.	1.73
I think that I would need the support of a technical person to be able to use this system.	4.18
I found the various functions in the system were well integrated.	1.90
I thought there was too much inconsistency in this system.	3.73
I would imagine that most people would learn to use this system very quickly.	1.45
I found the system very cumbersome to use.	3.63
I felt very confident using the system.	1.73
The system is accurate.	1.82
The system is unreliable.	3.90
The interaction with the system is unpredictable.	2.73
The system didn't always do what I expected.	2.73
The system is dependable.	2.09
The system makes few errors.	3.27
The interaction with the system is consistent.	2.09
The interaction with the system is efficient.	2.67

Mean score on 5 point Likert scale reported, (1 = strongly agree; 2=agree; 3=neither agree nor disagree; 4=disagree; 5=strongly disagree)

Table 7.

Self-efficacy to disclose HIV status before and after using Tough Talks VR program, Phase 83 pilot study (n=11)

I can disclose my HIV status before having sex, even to...	Mean Pre-test Score [SD]	Mean Post-test Score [SD]	p-value
A really hot new sex partner.	2.27 [1.351]	2.91 [1.451]	0.30
A really hot new sex partner who I think might be HIV negative.	2.45 [1.44]	2.36 [1.29]	0.88
A new sex partner when I'm really depressed or upset.	2.55 [1.37]	2.73 [1.10]	0.74
A partner who did not know I was positive the first time we had sex.	2.91 [1.64]	2.73 [1.35]	0.78
A sex partner who has not told me his HIV status.	2.09 [1.14]	2.18 [1.17]	0.85
A partner who may not want to have sex with me if they knew I was HIV positive.	2.82 [1.33]	2.82 [1.25]	1.00

Mean score on 5-point Likert scale reported, 1 = Definitely could disclose; 2 = Probably could disclose; 3 = Neither sure nor unsure; 4 = Probably could not disclose; 5 = Definitely could not disclose