NATIONAL SURVEY ON DRUG USE AND HEALTH:

INVESTIGATION INTO TEXT TO SPEECH SOFTWARE

DISCLAIMER

SAMHSA provides links to other Internet sites as a service to its users and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.

Substance Abuse and Mental Health Services Administration
Center for Behavioral Health Statistics and Quality
Rockville, Maryland 20857

April 2014

This page intentionally left blank

NATIONAL SURVEY ON DRUG USE AND HEALTH:

INVESTIGATION INTO TEXT TO SPEECH SOFTWARE

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.108.006.005

RTI Authors:

Hilary Zelko
Martin Meyer
Gilbert Rodriguez
David Roe
Patty LeBaron

Project Director:

David Hunter

SAMHSA Project Officer:

Peter Tice

For questions about this report, please email Peter.Tice@samhsa.hhs.gov.

Prepared for Substance Abuse and Mental Health Services Administration, Rockville, Maryland

Prepared by RTI International, Research Triangle Park, North Carolina

April 2014

Recommended Citation: Center for Behavioral Health Statistics and Quality. (2014). *National Survey on Drug Use and Health: Investigation into Text to Speech Software.* Substance Abuse and Mental Health Services Administration, Rockville, MD.

Acknowledgments

This publication was developed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality (CBHSQ), by RTI International, a trade name of Research Triangle Institute, Research Triangle Park, North Carolina, under Contract No. HHSS283201000003C. Contributors to this report from CBHSQ include Joel Kennet and Grace O'Neill. At RTI, significant contributors included Doug Currivan who reviewed the report, Richard Straw who copyedited the document, and Valerie Garner who provided production assistance.

Table of Contents

Chapter **Page**

List of Tables iv

1. Introduction 1

2. Initial Assessment of TTS Software Products 5

3. Phase 2 TTS Evaluation 7
   - 3.1 Methods 7
   - 3.2 Results 8
     - 3.2.1 Product 1: NeoSpeech 9
     - 3.2.2 Product 2: Microsoft 9
     - 3.2.3 Product 3: TextSpeech Pro 10
   - 3.3 Static Versus Dynamic Implementation 11

4. Costs and Impact on Work Processes 13

5. Conclusions and Recommendations 15

References 17

Appendix

A ACASI Questions Selected for TTS Prototypes 19

List of Tables

Table **Page**

3.1 Text to Speech (TTS) Prototypes....7

3.2 Product Averages for Evaluated Text to Speech (TTS) Prototypes....8

3.3 Gender of Voice Preference....8

3.4 Average Ratings on Quality Dimensions for NeoSpeech English Prototypes....9

3.5 Average Ratings on Quality Dimensions for NeoSpeech Spanish Prototypes....9

3.6 Average Ratings on Quality Dimensions for Microsoft English Prototypes....9

3.7 Average Ratings on Quality Dimensions for Microsoft Spanish Prototypes....10

3.8 Average Ratings on Quality Dimensions for TextSpeech Pro English Prototypes....10

4.1 Estimated Software Costs of Text to Speech (TTS) Products....13

5.1 Timeline for a Dynamic Text to Speech (TTS) Implementation Versus Current Approach....16

1. Introduction

Audio computer-assisted self-interviewing (ACASI) is a survey technology in which question text is simultaneously displayed on screen and read aloud to respondents. It has been adopted widely by survey researchers and has been used in a number of national Federal surveys because of its effectiveness in eliciting more accurate responses to highly sensitive or personal questions (Tourangeau & Smith, 1996) as well as for enabling participation among respondent populations with low literacy levels or sight impairments (Phillips, Edwards, & Dolbow, 2013). Typically, ACASI uses human voices recorded in the form of WAV or MP3 files that read the survey questions and response options on each screen. This requires a costly and labor-intensive effort both to record and edit high-quality audio and to programmatically integrate audio files with survey software (Phillips et al., 2013). This process can be especially difficult for questionnaires that dynamically generate question text using previous response fills or other changing components, such as dates, which need separate audio files that must be "stitched" together and can cause breaks or pauses in the flow of the audio. Another challenge for voice-recorded ACASI questionnaires is the reliance on a particular human voice that may change over time or become unavailable, thereby requiring the rerecording of all audio files with a replacement voice in order to ensure consistency within and across interviews. A potential solution to these challenges is the integration of text to speech (TTS) software, which uses a computer-generated voice to read text displayed on-screen. This text can be recorded in WAV file format (static) or run dynamically in conjunction with the survey instrument (dynamic), thereby eliminating the use of audio files altogether (Phillips et al., 2013). Compared with recording WAV files with a human voice, either approach has the potential to simplify the production of audio content associated with modifications to questionnaire text or the incorporation of new survey items.

In survey data collection, TTS is most commonly used in interactive voice response (IVR) or telephone audio computer-assisted self-interviewing (T-ACASI) modes. Surveys deployed in these modes are conducted by telephone, and the survey instructions and questions are delivered via prerecorded scripts (Couper, Singer, & Tourangeau, 2004). Using TTS systems to provide audio for ACASI in-person interviews is not yet as common as in telephone data collection modes. As Couper (2005, p. 488) noted, continued improvement to the quality of TTS technology "… opens the way to increased use of TTS systems for replacing the recording of interviewers [i.e., human voices]," including ACASI. Kraft and Taylor (2006) provided an early example of investigating TTS for use with an ACASI protocol in which they suggested that the use of TTS provided greater flexibility in making edits and changes and reduced development time and costs while providing respondents with an impersonal solicitor for sensitive questions. In 2009, the Substance Abuse and Mental Health Services Administration (SAMHSA) and staff at RTI International investigated the use of TTS for replacing the recorded human voice for the ACASI sections of the National Survey on Drug Use and Health (NSDUH) interview (LeBaron, Meyer, & Rodriguez, 2009). This investigation indicated that the TTS systems available at that time had several potential benefits, but the effort required to customize the off-the-shelf pronunciations in order to meet quality standards would be significant. However, in recent years, advances in TTS synthesis technology have been enabling the generation of more realistic, accurate, and human-sounding "voices" in multiple languages that may be customized with

respect to pacing, volume, pitch, and pronunciation. Concurrent with these advances, the use of computer-generated voices has gone from a few, highly technical applications to the point where many people encounter synthesized speech regularly in their day-to-day lives, and are becoming more accustomed to it.

In response to this improved technology, survey researchers have begun experimenting with implementing TTS. In September 2011, the developers of the National Survey of Family Growth (NSFG) replaced human voice recordings with static TTS recordings for the ACASI portion of the Cycle 8 interview, and they compared data from Cycle 8 with Cycle 7 data that used human voice recordings (Couper, Kirgis, Buageila, & Berglund, 2012). In addition to supporting past literature on the overall cost and time efficiencies provided by TTS in preparation for data collection, the NSFG researchers found that more respondents were making use of ACASI, but taking less time (in terms of overall interview time) to do so, suggesting that smaller audio files, less downtime or "space" between files and fills, and the pace of TTS led to better overall interview efficiency (and, one could assume, reduced burden) while in the field. In 2012, the Population Assessment of Tobacco and Health (PATH) study also turned to TTS for a field test. Using a dynamic implementation mode that uses the TTS engine to read question text in real time and eliminates the use of prerecorded audio WAV files altogether, Phillips et al. (2013) supported the aforementioned research on the efficiency of using TTS, especially when it comes to eliminating the sheer size and volume of files that must be "stitched" together in order to provide a smooth delivery of audio. They noted that in this mode TTS consumes far less resources (server space, storage, etc.) than live voice recordings.[1]

It is in this context that SAMHSA and RTI have undertaken a follow-up investigation of TTS software to update prior research, assess which of the commercially available TTS products are potentially viable for use on NSDUH, and determine the pros and cons of utilizing each on NSDUH. The investigation has been conducted in three phases: Phase 1 focused on researching different TTS systems and identifying products suitable for further evaluation; Phase 2 consisted of an evaluation of selected products; and Phase 3 focused on the development of this report and a recommendation regarding which of the evaluated products is most promising for NSDUH's ACASI modules.

During Phase 1 of this investigation, RTI conducted a broad review of the survey research methods literature and the TTS software landscape to identify research on TTS implementation for ACASI and specific TTS products that could potentially be implemented on NSDUH. Because many TTS products offer both female and male voices, the review of the survey research methods literature included studies that examined how the gender of an ACASI voice might affect survey response.

The literature remains inconclusive regarding the impact of the gender of the voice on survey response. Outside of the survey research industry, scholars and technology experts have suggested that computer voices are mostly female due to biology (people finding female voices more pleasing than male voices) or history (e.g., telephone operators) (Griggs, 2011). That said, gender experts have suggested that gender stereotypes, if present in an individual, can in fact

[1] The PATH study did not report any results associated with the impact on survey response because a full implementation of TTS had not been completed at that time.

extend to machines, suggesting that voice selection (in terms of gender) could be highly consequential (Nass, Moon, & Greene, 1997).

In the survey research literature, information is somewhat scarce, although Dykema, Diloreto, Price, White, and Schaeffer (2012) presented an excellent review of the literature at the beginning of their investigation into ACASI gender-of-voice effects on the reporting of sensitive behaviors among young adults. They explained that, depending on a question's topic, respondents may refer to gender-based stereotypes, conversational norms, or identities when responding (Tannen, 1996; see also Schaeffer, 2000). According to self-disclosure theory, individuals are expected to be more honest and disclose more to someone they trust and with whom they feel comfortable (Jourard, 1971). Insofar as respondents hold stereotypes that women are more sympathetic (Pollner, 1998) or nonjudgmental (Nass, Robles, Heenan, Bienstock, & Treinen, 2003), respondents may disclose or report higher levels of sensitive behaviors to female interviewers (Dindia & Allen, 1992). In contrast to self-disclosure theory, other researchers have offered "explanations of exaggeration," which hold that higher levels of reporting may be less valid. For example, the "macho hypothesis" of Catania et al. (1996, p. 371) explains the higher levels of some sexual behaviors that males report to male interviewers as an effort to seem more virile and manly. In a related vein but predicting a different outcome, Weisel (2002, p. 102) argued in her study of contemporary gangs that "a female interviewer may [have] inadvertently encourage[d] male interviewees to put on a macho bravado and exaggerate some points." However, Dykema et al. (2012) found higher levels of engagement in sensitive behaviors and more consistent reporting among males when responding to a female voice. They did not find any evidence that female respondents were influenced by the voice's gender.

Couper et al. (2004) also found no significant difference between male and female voices on the reporting of sensitive behaviors or gender attitudes in a telephone survey. They compared question administration by live telephone interviewers, an IVR system that used human voice recordings, and an IVR system that used synthesized speech generated from TTS software. Equal numbers of male and female respondents were enrolled in the study. Although the authors expected to find gender-of-voice effects for gender attitudes and some sensitive questions (i.e., sexual behavior, weight) in both the live interview and IVR conditions, surprisingly, they found no significant differences due to the gender of the voices used. However, they did find consistent effects for differences in responses to the live interviewers as compared with the automated IVR system, which consistently revealed greater disclosure of sensitive behaviors in the IVR conditions. This finding is consistent with the research methods literature that suggests that ACASI facilitates more accurate and candid reporting of sensitive behaviors, though it does not indicate that the gender of the voice used in ACASI always has a significant impact on survey responses. One other study, conducted exclusively with male respondents, directly tested the effect of the ACASI voice's gender on survey reports; Fahrney, Uhrig, and Kuo (2010) explored the impact of a male versus female voice on reports of sexual activity among men who have sex with men (MSM). Their results were consistent with more accurate reporting among the males who heard questions read by a female voice. Although Dykema et al. (2012) concluded that a female voice should be the convention for ACASI studies, they also noted that more experimentation and investigation are required in order to ensure the elimination of underreporting across specialized topics and populations.

2. Initial Assessment of TTS Software Products

As part of the first phase of this evaluation, RTI conducted a broad review of the text to speech (TTS) landscape to identify software products that could be considered for further hands-on assessment in Phase 2. A wide variety of TTS software products is currently available, ranging from small Web plug-ins or mobile apps (i.e., applications), to stand-alone desktop software, to server-based solutions for streaming speech output. The sound and quality of the speech output depend largely on the speech synthesizer software that generates the "voice" and on the methods available to vary the pacing, volume, emphasis, or pronunciation of terms. Products vary with respect to the synthesizer technology they use and the functionality provided for customization. For example, some products provide a "studio" editing interface that allows for easier manual editing and manipulation of the pronunciation, volume, and pacing of the speech output, while others require customization using programming code, such as Speech Synthesis Markup Language (SSML). Currently, Microsoft, Apple, and Android each has a TTS synthesizer built into its operating system. In addition, Microsoft's Speech Platform Software Development Kit (SDK) allows developers to build applications that interact with the synthesizer programmatically. Other commercial providers, such as AT&T Natural Voices, Acapela, and NeoSpeech, have developed their own TTS synthesizers that can run on the Windows platform and be integrated with a variety of applications. They may also provide editing or "studio" software that can be used for customizing the speech output. Other products typically provide a user interface for inputting text or selecting documents/Web pages to be read and use either the operating system's built-in synthesizer or one of the commercial third-party synthesizers, such as AT&T or NeoSpeech.

In order to identify feasible options for the Phase 2 evaluation, RTI staff conducted online research and contacted a variety of vendors to learn about different TTS products. The most important factors in the selection of products are the sound and quality of the voice, the ability to customize speech output (pronunciation/pacing), and the ability to run on a laptop in an offline environment. Also, the search was limited to vendors that have a track record in the industry and are likely to remain key players with well-supported products. As requested by the Substance Abuse and Mental Health Services Administration (SAMHSA), the research began with the three products evaluated as part of the 2009 TTS investigation (i.e., Acapela, AT&T Natural Voices, and NeoSpeech) to determine whether they remained feasible options for the Phase 2 evaluations. These products have continued to evolve since 2009 and are still considered among the top commercially available TTS products. RTI staff examined software features, listened to online demos of the voices, and spoke with company representatives to discuss the software's capability to provide both a dynamic speech solution and a static, prerecorded audio file-based solution. The conclusion was that all three products remain viable options for further investigation. Next, RTI staff extended the search to include other products that have been used by other survey groups or have come on the market or gained prominence since 2009. TextSpeech Pro emerged as an additional product for consideration because it was used on both the National Survey of Family Growth (NSFG) and Population Assessment of Tobacco and Health (PATH) surveys (see the discussion in Chapter 1).

In addition to the four products mentioned above, RTI staff researched a wide variety of new TTS products. Several turned out to be less suitable products with limited capabilities, marketed mainly for personal use, so the team focused on seven mainstream and commercially available TTS packages, each of which supports integration with other applications, as follows:

- TextSpeech Pro,
- AT&T Natural Voices,
- NeoSpeech,
- Microsoft Speech Platform,
- Acapela,
- Ivona, and
- Loquendo.

The Phase 2 evaluation plan required an assessment of the two core modes for implementing TTS software in the National Survey on Drug Use and Health (NSDUH) audio computer-assisted self-interviewing (ACASI) module: *static* and *dynamic* implementations. "Static" implementation uses TTS software to record all of the audio files needed by the ACASI instrument. These recordings are saved as WAV files, stored on the laptop, and are played by the ACASI program as the respondent moves through the interview. This is similar to NSDUH's current ACASI approach, with the primary difference being that the TTS "voice" is recorded rather than a human voice. "Dynamic" implementation installs the TTS engine on each laptop, and the ACASI program interfaces with the TTS engine to construct the audio in real time as each question appears on screen.

The features of the TTS packages were examined with respect to both implementation modes. The dynamic implementation required a Windows-based desktop SDK or an Application Programming Interface (API) in order to interact with NSDUH's Blaise survey software. The static audio file-based approach required that TTS speech could be output as WAV audio files. In summary, the TTS software features required to support the NSDUH's needs are as follows:

- for dynamic implementation, Windows-compatible SDK or API to allow integration with the NSDUH interview software and the ability to run in an offline environment;
- for static implementation, the ability to output speech to WAV file format;
- customizable pronunciation and extensible dictionaries to accommodate specialized terminology;
- graphical user interface (GUI) that enables customization of volume, pausing, and speaking rate;
- availability of male and female voices; and
- availability of English and Spanish voices.

Three of the seven products listed above satisfied the requirements: TextSpeech Pro, NeoSpeech, AT&T Natural Voices. The Microsoft Speech Platform met most requirements, but did not include any male voices. The requirements above excluded three of the seven TTS products listed above: Acapela, Ivona, and Loquendo.

3. Phase 2 TTS Evaluation

This chapter describes the Phase 2 evaluation methods and results. Additionally, RTI's technical team evaluated the advantages and disadvantages of the static and dynamic implementation modes, which are described in Section 3.3.

3.1 Methods

After discussions with the National Survey on Drug Use and Health (NSDUH) team at the Substance Abuse and Mental Health Services Administration (SAMHSA), three products were selected for purchase: TextSpeech Pro, NeoSpeech, and AT&T Natural Voices.[2] Shortly before the purchase was made, AT&T unexpectedly pulled their Natural Voices product off the market without providing any information on possible future availability. As a result, SAMHSA and RTI staff decided to eliminate the AT&T product from further consideration, and the Microsoft Speech Platform (which met most of the requirements identified in Phase 1) was used to replace the AT&T product in the Phase 2 evaluation.

An initial set of prototypes using a subset of 12 audio computer-assisted self-interviewing (ACASI) questions was developed using the static and dynamic modes, English and Spanish languages, and male and female voices. The computer-assisted interviewing (CAI) questions were selected to include a mix of specialized terms, such as the names of alcoholic beverages or prescription drug names, long sections of text, and question fills. These questions are listed in Appendix A. Because many of the default pronunciations, particularly for specialized terms and prescription drug names, were incorrect, the technical team worked to fine-tune and enhance these pronunciations and customize the pausing, speed, and pitch of voices to maximize the probability that respondents could understand the voice. The team spent approximately 10 minutes per screen to enhance the quality, which provided a realistic test of the quality that could be achieved in a full-scale implementation.

A total of 12 prototypes—8 in English and 4 in Spanish—were developed and are listed in Table 3.1. These included male, female, and Spanish versions that were available from each product. Although originally both female and male voices were required for testing, no male Spanish voices were available for any of the products, so only female Spanish voices were tested. Also, no male English voice was available for the Microsoft Speech Platform.

Table 3.1 Text to Speech (TTS) Prototypes

English Prototypes	Spanish Prototypes
1. NeoSpeech Male Static	1. NeoSpeech Female Static
2. NeoSpeech Female Static	2. Microsoft Female Static
3. Microsoft Female Static	3. NeoSpeech Female Dynamic
4. TextSpeechPro Female Static	4. Microsoft Female Dynamic
5. TextSpeechPro Male Static	
6. Microsoft Female Dynamic	
7. NeoSpeech Female Dynamic	
8. NeoSpeech Male Dynamic	

[2] The Microsoft Speech Platform product was not included initially because it did not need to be purchased and an initial review indicated that it did not meet all of the requirements. Prior to Phase 2, a newer and higher quality voice option offered by Microsoft was identified and included in the Phase 2 evaluation.

A team of seven evaluators (five English speakers and two Spanish speakers), including one male and six females, with experience in survey research were recruited to assess the prototypes. Both Spanish-speaking evaluators had experience in designing Spanish-language surveys.

Each evaluator was assigned to assess four prototypes in a specific order. Assignments were made so that evaluators were presented with a mix of static, dynamic, male, and female prototypes in different orders to minimize any potential bias due to the order in which evaluators heard the prototypes. Each prototype was assessed by at least two evaluators. The evaluators were instructed to navigate through the CAI prototype and rate each screen along the following dimensions using a rating scale from 1 to 5: clarity, inflection, tone, humanness, and pace. The points on the scale were labeled as follows:

Numeric Value	Label
1	Poor
2	Fair
3	Good
4	Very Good
5	Excellent

Ratings were then averaged across all screens, and the evaluators generated an average product score for each prototype. After completing the ratings, evaluators were asked to report a gender preference and which, if any, of the prototypes they preferred.

3.2 Results

This section presents the results of the evaluation. Ratings on all screens for each dimension were averaged across all evaluators to provide an overall product average for each prototype (see Table 3.2). The gender of voice preferences are displayed in Table 3.3. The detailed results for each product are displayed in the sections that follow.

Table 3.2 Product Averages for Evaluated Text to Speech (TTS) Prototypes

Product	English Static Female	English Dynamic Female	English Static Male	English Dynamic Male	Spanish Static Female	Spanish Dynamic Female
NeoSpeech	2.6	2.9	2.8	3.0	2.1	2.5
Microsoft	2.6	2.9	n/a	n/a	2.1	2.8
TextSpeech Pro	2.1	n/a	2.0	n/a	n/a	n/a

n/a = not available.

When asked if they had a preference for a male or female voice, most evaluators were indifferent or preferred female, as shown in Table 3.3.

Table 3.3 Gender of Voice Preference

Gender of Voice Preference	Male	Female	Indifferent
Number of Evaluators	1	2	2

3.2.1 Product 1: NeoSpeech

The audio quality ratings for the NeoSpeech English prototypes are shown in Table 3.4, while ratings for the Spanish prototypes are displayed in Table 3.5. The product average scores for NeoSpeech's female English voice tied Microsoft's female English voice in both the static and dynamic versions (see Section 3.2.2 for the Microsoft product results). NeoSpeech's static prototypes were ranked higher than both TextSpeech Pro's female and male prototypes (see Section 3.2.3 for the TextSpeech Pro product results). The dynamic implementations were higher than the static versions and were the highest of all female prototypes. NeoSpeech's male English voice was ranked slightly higher than the female voice. Despite these higher rankings, only one English evaluator preferred a male voice over a female one.

Table 3.4 Average Ratings on Quality Dimensions for NeoSpeech English Prototypes

NeoSpeech English				
Dimension	**Static Female**	**Dynamic Female**	**Static Male**	**Dynamic Male**
Clarity	2.8	3.3	2.9	3.6
Inflection	2.5	2.7	2.6	2.8
Tone	2.9	2.9	3.4	2.9
Humanness	2.0	2.5	2.2	2.6
Pace	2.8	3.2	3.0	3.1
Product Average	*2.6*	*2.9*	*2.8*	*3.0*

Table 3.5 Average Ratings on Quality Dimensions for NeoSpeech Spanish Prototypes

NeoSpeech Spanish		
Dimension	**Static Female**	**Dynamic Female**
Clarity	2.4	2.5
Inflection	1.9	2.3
Tone	1.9	2.4
Humanness	2.1	2.3
Pace	2.3	3.1
Product Average	*2.1*	*2.5*

The product average for NeoSpeech's female Spanish voice was higher for the dynamic version than the static version, but it was lower than the product average for Microsoft's female Spanish voice. No male Spanish voices were available from NeoSpeech or Microsoft.

3.2.2 Product 2: Microsoft

The audio quality ratings for the Microsoft English prototypes are displayed in Table 3.6, while the ratings for the Spanish prototypes are shown in Table 3.7. Microsoft's female English voice tied NeoSpeech's female English voice for the highest ranking of the female voices and outranked the TextSpeech Pro female prototype by one-half rating point. As with NeoSpeech, the product average for the dynamic version was higher than the static version. No male English voices were available with the Microsoft Speech Platform at the time of evaluation.

Table 3.6 Average Ratings on Quality Dimensions for Microsoft English Prototypes

Microsoft English		
Dimension	**Static Female**	**Dynamic Female**
Clarity	3.5	3.9
Inflection	2.2	2.8
Tone	2.7	2.9
Humanness	1.7	2.5
Pace	2.8	2.7
Product Average	*2.6*	*2.9*

Table 3.7 Average Ratings on Quality Dimensions for Microsoft Spanish Prototypes

Microsoft Spanish		
Dimension	**Static Female**	**Dynamic Female**
Clarity	2.1	2.8
Inflection	2.0	2.7
Tone	2.0	2.8
Humanness	2.4	2.8
Pace	2.1	3.0
Product Average	*2.1*	*2.8*

Microsoft's female Spanish voice in the dynamic version was ranked the highest of all Spanish voices. The static version had a product average equivalent to NeoSpeech's female Spanish static version.

3.2.3 Product 3: TextSpeech Pro

The product averages for the TextSpeech Pro female and male English prototypes were the lowest of all English prototypes (see Table 3.8). They scored lower on almost every dimension than both NeoSpeech's female and male static prototypes as well as Microsoft's female static prototype. A dynamic version of TextSpeech Pro was unavailable, and no Spanish voices were available with the TextSpeech Pro software.

Table 3.8 Average Ratings on Quality Dimensions for TextSpeech Pro English Prototypes

TextSpeech Pro English		
Dimension	**Static Female**	**Static Male**
Clarity	2.7	2.4
Inflection	1.8	1.8
Tone	1.9	2.0
Humanness	1.5	1.4
Pace	2.8	2.5
Product Average	*2.1*	*2.0*

The product averages for the TextSpeech Pro female and male English prototypes were the lowest of all English prototypes (see Table 3.8). They scored lower on almost every dimension than both NeoSpeech's female and male static prototypes as well as Microsoft's

female static prototype. A dynamic version of TextSpeech Pro was unavailable, and no Spanish voices were available with the TextSpeech Pro software.

3.3 Static Versus Dynamic Implementation

As part of the activity associated with developing the prototypes, NSDUH's technical team at RTI also assessed requirements and tradeoffs associated with the static versus dynamic TTS implementations. This section presents a summary of the key advantages and disadvantages of each approach. Overall, the team concluded that the advantages of the dynamic approach, which allows greater flexibility, efficiency, and audio quality, outweigh those of a static approach. In addition, the dynamic prototypes were consistently rated higher than the static prototypes.

Using the static TTS approach to prerecord audio files would have a number of advantages associated with eliminating the process for human voice recording and minimizing the impact on the development process:

- Audio files could be generated in a batch from text descriptions of the questions. This would require significantly less time than human voice recording and would not require the use of a recording studio.
- Also, in a static TTS implementation, the Blaise program on the laptop would remain largely the same as the current version, with only minimal changes required.
- Installing TTS software on the laptops would be unnecessary, and the configuration of NSDUH's field laptops would remain largely the same, with audio files installed locally on each laptop.
- Testing would be similar to the current approach and somewhat easier than a dynamic implementation because no TTS software would be required to listen to audio files, which could be played on any computer or even a mobile device.

One of the main drawbacks of the static approach is that audio files must still be "stitched" together, which would not eliminate some of the breaks or pauses that can occur when multiple files must be played for a single question. Also, the static implementation would require the production of approximately 10,000 audio files (5,000 for ACASI in English and 5,000 for ACASI in Spanish) to replace the human voice recordings currently used on NSDUH. This work would require the purchase or development of software to generate and edit the audio files on RTI development computers, and it would also require the development of a database for storing and maintaining all of the audio files and text scripts for each audio segment.

A dynamic implementation offers a number of advantages that are largely associated with eliminating the use of audio files:

- Audio files would not be prerecorded, thereby eliminating the audio file production process and the effort associated with editing, prerecording, storing, and maintaining a database of audio files.
- The dynamic approach also offers a smoother audio delivery on questions that have fills because breaks or pauses that occur when audio files have to be "stitched" together would no longer be present.

- The elimination of audio files would simplify the process of distributing patches to in-field laptops because there would be no audio files to distribute as part of a patch. This would make patches smaller and easier to deploy, providing increased flexibility to make needed changes to the questionnaire during ongoing data collection.
- There would be no need to maintain audio files and scripts used for recording.

Among the disadvantages of a dynamic implementation is that more sophisticated programming is necessary, which requires familiarity with specific tools and features of the TTS product and the Windows Speech Application Programming Interface (API). Although a number of changes would be required in the Blaise programming code to implement a dynamic approach, the experience acquired during the development of the prototypes allowed the team to establish methods that could be implemented in the entire set of ACASI modules fairly easily. Additionally, the dynamic approach would tend to make the interview software testing process somewhat more cumbersome because each testing laptop would require installation and configuration of TTS software, whereas in a static implementation, there is a finite set of audio files that can be easily shared and played on any laptop. Also of interest is the fact that the dynamic implementation requires that TTS software is installed and configured on every field laptop, which could potentially introduce complications associated with software licensing agreements. In particular, the NeoSpeech TTS product requires each laptop to be configured with a unique license activation key, permanently tied to the MAC address (the hardware ID number) of that laptop. As such, a dynamic implementation based on NeoSpeech severely complicates the duplication and maintenance of NSDUH field laptops, and it requires that a NeoSpeech license be purchased for every laptop. It is certainly worth noting here that the Microsoft TTS product, which is free, does not impose any such licensing requirements or restrictions. Therefore, a dynamic implementation based on the Microsoft TTS product would not have the same drawbacks associated with a dynamic implementation based on NeoSpeech.

Finally, it should be noted that regardless of mode, TTS implementation in the first year would require a significant effort with respect to customizing pronunciations and testing the audio components. Many of the default pronunciations for specialized terms and drug names would need to be corrected using alternate, phonetically based spellings. Furthermore, even with customization, there are likely to be some terms or phrases for which the pronunciations are suboptimal, which is true for both the static and dynamic implementations. The technical team estimated that the level of effort to customize and test the audio components would be similar in either the static or dynamic approaches.

4. Costs and Impact on Work Processes

Cost is a key factor in the adoption of any new technology. The implementation of text to speech (TTS) software offers the potential for work process efficiencies and long-term cost savings associated with the development and update of the National Survey on Drug Use and Health (NSDUH) audio computer-assisted self-interviewing (ACASI) program. Startup and ongoing costs are associated with the purchase and use of the TTS software and the level of effort required to implement the new technology. This chapter discusses both the software costs and the impact on the level of effort in key work processes involved in TTS implementation.

Each of the products evaluated in Phase 2 is summarized in Table 4.1. Estimated costs associated with the purchase and ongoing use of the product are provided in both static and dynamic implementations. Licensing requirements are different for each product. NeoSpeech requires an annual license agreement with the purchase of their products, while Microsoft does not. TextSpeech Pro licenses are bundled in the purchase prices of the software on a one-time, per laptop basis, and no renewal is required.

Table 4.1 Estimated Software Costs of Text to Speech (TTS) Products

TTS Product	License Required	Cost: Static Implementation	Cost: Dynamic Implementation
NeoSpeech	Yes, annual renewal	$5,300/year (700 laptops)	$9,800/year (700 laptops)
Microsoft	No	$0	$0
TextSpeech Pro	Yes, one time	$1,500 (10 development computers)	NA[1]

NA = not applicable.

[1] Although RTI staff were initially told that a Software Development Kit (SDK) was available for implementing a dynamic version, only a demo version could be obtained that appeared to be last updated in 2009. RTI staff were later told that the voice provider (i.e., NeoSpeech, AT&T) would need to contacted to obtain the SDK needed to implement a dynamic version.

RTI staff also reviewed key work processes involved in NSDUH's current ACASI development effort that would be affected by TTS implementation to assess whether an increase, decrease, or similar level of effort would be required. These work processes are in four areas:

1. production of audio files for ACASI,
2. customization/editing and testing of audio components,
3. ACASI programming, and
4. field laptop configuration and mass duplication.

The production of audio files in the current approach involves activities associated with human voice recording using a recording studio and sound engineer and the maintenance of a database for storing and maintaining audio files. A reduced level of effort and greater efficiency over time would be expected in a static implementation because audio files could be generated

from text scripts in batch using purchased or developed software, thereby eliminating costs associated with human voice recording and rerecording. Clearly, in the first year of implementation, a learning curve would be associated with using the new software, implementing new processes, and generating scripts for approximately 10,000 new audio files. In a dynamic implementation, this process would be eliminated altogether.

The second area (i.e., customization/editing and testing of audio components) will require the greatest increase in effort, regardless of implementation mode, especially in the first year of implementation. The RTI staff's experience in developing the TTS prototypes revealed that default pronunciations for specialized terms (e.g., alcohol types, prescription drugs, and health conditions) had to be customized using alternate phonetic spellings. The speaking rate, pitch, and volume also were modified to improve the quality. As a result, it is expected that a greater number of audio components would require editing and quality control checks than in the current approach where pronunciations, pacing, and pitch are more easily adjusted by the human voice in the studio at the time of initial recording. Therefore, TTS implementation is expected to require an increased level of effort associated with the editing of pronunciations compared with the current approach. However, the same level of effort for customization/editing and testing is expected, regardless of whether a static or dynamic implementation is adopted.

In the third area, the ACASI programming effort, a static implementation would require essentially the same effort as in the current approach, and minimal changes would be required in the Blaise code to accommodate the TTS audio files. A dynamic implementation would require more complex programming to integrate the TTS engine with the Blaise software. However, because RTI staff developed the programming approach and an initial set of code to achieve this while building the prototypes, a small increase is anticipated in programming effort to implement for the full set of ACASI modules.

Similarly, the process for field laptop configuration and mass duplication (the fourth area) would be similar to the current effort in a static implementation and somewhat more complicated for a dynamic implementation, which requires TTS software to be installed and configured on each field laptop. As noted earlier, this process is significantly more complex if the NeoSpeech product is deployed because of its hardware-based annual licensing structure, which would complicate not only the initial deployment, but also annual updates and ongoing replacements of field laptops. Because the Microsoft Speech Platform does not require a license, the configuration process would be significantly simplified.

5. Conclusions and Recommendations

Implementing text to speech (TTS) technology in the audio computer-assisted self-interviewing (ACASI) module of the National Survey on Drug Use and Health (NSDUH) interview offers an opportunity for work process efficiencies and cost savings in NSDUH's ACASI software development. Although TTS software cannot match the audio quality of human voice recording, the evaluation presented in this report, along with the experience of both the National Survey of Family Growth (NSFG) and the Population Assessment of Tobacco and Health (PATH) survey, indicate that it is of sufficient quality to serve as a replacement for human voice recordings in ACASI. The goal of implementation is to create audio recordings that can be easily understood by respondents of all ages in both English and Spanish. None of the evaluators had difficulty understanding any words or phrases produced by the TTS voices. A significant advantage of TTS, regardless of implementation mode, is the elimination of the effort and costs associated with the use of a recording studio and human voices for generating ACASI audio. It also eliminates the reliance on a particular human voice, which may change or become unavailable over time.

The evaluation results indicate that a dynamic implementation offers a higher quality audio experience than the static implementation largely because of the elimination of audio files. The dynamic approach also eliminates processes for producing, maintaining, and storing audio files, and it would simplify the process for updating or modifying NSDUH's ACASI modules. The dynamic approach outperformed the static approach in the evaluation, and all of the evaluators indicated that they preferred the dynamic prototypes. Therefore, if TTS software is adopted for use with NSDUH, a dynamic implementation mode should be pursued over a static mode. With respect to the TTS products evaluated, Microsoft's Speech Platform is recommended. Considering the evaluated TTS products in both English and Spanish, the Microsoft product was ranked the highest by the evaluation team. Also, the Microsoft product offers a significant advantage over NeoSpeech because it is freely available and requires no licensing agreement or user fees.

A dynamic implementation of TTS software for the 2015 NSDUH would have several implications for the questionnaire development schedule. First, time allocated for recording and internal testing of WAV files would be eliminated. Additional time would be required for editing and reviewing the TTS audio components of the ACASI modules. Also, additional time would be required to test the complete set of ACASI modules. If the Microsoft product is used, the TTS software could be bundled with the computer-assisted interviewing (CAI) master update image and configured programmatically, which could be completed in the same timeframe as the current approach. As a result, the CAI instrument could be finalized in early October 2014, following the typical schedule, but creation and verification of audio components should begin as soon as possible.

Table 5.1 displays an estimated timeline for a dynamic TTS implementation compared with the current approach.

Table 5.1 Timeline for a Dynamic Text to Speech (TTS) Implementation Versus Current Approach

Timeline	TTS Approach		Current Approach	
Questionnaire Development	**Begin Date**	**End Date**	**Begin Date**	**End Date**
Submit CAI Specifications to SAMHSA with OMB Package	2/14/14	2/14/14	2/14/14	2/14/14
Submit Initial CAI Specifications to Programming Team	2/14/14	2/14/14	3/3/14	3/3/14
Translate Initial Specifications into Spanish	2/17/14	3/28/14	3/4/14	4/1/14
Initial Review and Editing of TTS Pronunciations for the ACASI Modules	1/28/14	2/24/14	NA	NA
Secondary Review of Edited TTS Pronunciations	2/26/14	3/24/14	NA	NA
Program and Test TTS Audio in Initial CAI Questionnaire	2/17/14	6/23/14	NA	NA
Send ACASI Modules on a Flow Basis to SAMHSA for Review	4/1/14	6/26/14	NA	NA
Receive SAMHSA Feedback on TTS Audio	4/30/14	6/26/14	NA	NA
Screen Text Lockdown	6/26/14	6/26/14	7/14/14	7/14/14
Incorporate TTS Audio into 2015 CAI Instrument and Test	7/1/14	7/29/14	NA	NA
Send Instrument to SAMHSA for Testing	7/30/14	7/30/14	8/6/14	8/30/14
Receive Feedback from SAMHSA	7/30/14	8/18/14	9/1/14	9/15/14
Final Testing of TTS Instrument	8/18/14	9/2/14	NA	NA
Send 2015 CAI Instrument with TTS Audio to SAMHSA	9/3/14	9/3/14	NA	NA
Receive Final Feedback from SAMHSA on TTS Instrument	9/4/14	9/24/14	NA	NA
Final Questionnaire Program Lockdown	9/25/14	10/1/14	9/15/14	10/1/14
Send Final Instrument to SAMHSA	10/15/14	10/15/14	10/15/14	10/15/14

ACASI = audio computer-assisted self-interviewing; CAI = computer-assisted interviewing; NA = not applicable; OMB = Office of Management and Budget; SAMHSA = Substance Abuse and Mental Health Services Administration; TTS = text to speech.

References

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., & Coates, T. J. (1996). Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly, 60*, 345-375. doi:10.1086/297758

Couper, M. P. (2005). Technology trends in survey data collection. *Social Science Computer Review, 23*, 486-501. doi:10.1177/0894439305278972

Couper, M. P., Kirgis, N., Buageila, S., & Berglund, P. (2012). *Using text-to-speech (TTS) for audio-CASI*. Presented at the American Association for Public Opinion Research (AAPOR) 67th annual conference, Orlando, FL.

Couper, M. P., Singer, E., & Tourangeau, R. (2004). Does voice matter? An interactive voice response (IVR) experiment. *Journal of Official Statistics*, *20*, 551-570.

Dindia, K., & Allen, M. (1992). Sex differences in self-disclosure: A meta-analysis. *Psychological Bulletin, 112*, 106-124. doi:10.1037/0033-2909.112.1.106

Dykema, J., Diloreto, K., Price, J. L., White, E., & Schaeffer, N. C. (2012). ACASI gender-of-interviewer voice effects on reports to questions about sensitive behaviors among young adults. *Public Opinion Quarterly*, 76, 311-325. doi:10.1093/poq/nfs021

Fahrney, K. M., Uhrig, J., & Kuo, T.-M. (2010, April). *Gender-of-voice effects in an ACASI study of same-sex behavior* (RTI Press Methods Report MR-0017-100). Retrieved from http://www.rti.org/publications/rtipress.cfm?pubid=14766

Griggs, B. (2011, October 21). *Why computer voices are mostly female*. Retrieved from http://www.cnn.com/2011/10/21/tech/innovation/female-computer-voices/

Jourard, S. M. (1971). *Self-disclosure: An experimental analysis of the transparent self*. New York, NY: Wiley-Interscience.

Kraft, J., & Taylor, W. (2006, May). *Text-to-speech application in audio CASI: Evaluation of implementation and deployment*. Presented at the International Field Directors & Technologies Conference, Montreal, Quebec, CA.

LeBaron, P., Meyer, M., & Rodriguez, G. (2009, July). *2009 National Survey on Drug Use and Health: Text to speech investigation final report* (RTI 0209009486.023, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-2004-00022). Research Triangle Park, NC: RTI International.

Nass, C., Moon, Y., & Greene, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology, 27*, 864-876. doi:10.1111/j.1559-1816.1997.tb00275.x

Nass, C., Robles, E., Heenan, C., Bienstock, H., & Treinen, M. (2003). Speech-based disclosure systems: Effects of modality, gender of prompt, and gender of user. *International Journal of Speech Technology,* 6, 113-121. doi:10.1023/A:1022378312670

Phillips, J., Edwards, B., & Dolbow, E. (2013). *Using text-to-speech software for ACASI*. Presented at the Federal CASIC Workshops, Washington, DC.

Pollner, M. (1998). The effects of interviewer gender in mental health interviews. *Journal of Nervous and Mental Disease, 186*, 369-373.

Schaeffer, N. C. (2000). Asking questions about threatening topics: A selective overview. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 105-121). Mahwah, NJ: Lawrence Erlbaum Associates.

Tannen, D. (1996). *Gender and discourse*. New York, NY: Oxford University Press.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60*, 275-304.

Weisel, D. L. (2002). *Contemporary gangs: An organizational analysis*. New York, NY: LFB Scholarly Publishing.

Appendix A: ACASI Questions Selected for TTS Prototypes

A.1. Summary of Proposed CAI Items for Prototypes

The full question text and response options from the 2013 Dress Rehearsal computer-assisted interviewing (CAI) specifications are provided in Section A.2.

Variable Name	Description
1. Card 3a	Long list of types of alcoholic beverages
2. AL 01	Asks whether R has ever drank alcohol
3. HALINTRO	Introduction to hallucinogen module – lists various hallucinogens
4. INHINTRO	Introduction to inhalant module – lists various inhalants
5. TR03	Asks about past use of specific tranquilizers
6. TR05	Asks about past use of other tranquilizers
7. ST04	Asks about past use of specific stimulants
8. SV03	Asks about past use of specific sedatives
9. PRINTROYR2	Pre-fills a previous response
10. HLTH25	Asks whether R has had a specific set of health conditions
11. AD19	Includes pre-fills from previous response about mood problems
12. INTROINC	Asks about family income, includes fills and wording changes depending on family relationships and whether proxy is answering

R = respondent.

A.2. Selected Items from CAI Specifications – Full Question Text and Response Options (2013 Dress Rehearsal CAI Specifications)

A. English Versions

1. CARD3a Types of Alcoholic Beverages

Beer
Regular beer
Lite or light beer
Low-alcohol (LA) beer
Malt liquor
Ale
Stout
Lager

Wine
Red, white, blush wine
Wine coolers
Champagne
Sherry
Homemade wines, such as muscadine, scuppernong, or fruit wines
Fortified wines, such as Cisco

Liquor
Bourbon
Gin
Rum
Scotch
Tequila
Vodka
Homemade liquor, such as moonshine

Liqueurs, Cordials, and Brandy
Brandy
Cassis
Cognac
Creme de menthe
Drambuie
Grand Marnier
Kahlua
Port
Schnapps
Tia Maria
Triple sec
Vermouth

Mixed Drinks and Cocktails
Bloody Mary
Bourbon and water
Daiquiri
Gin and tonic
Manhattan
Margarita
Martini
Piña colada
Rob Roy
Rum and cola
Scotch and soda
Whiskey sour

Press [ENTER] to continue.

2. AL01

Have you **ever**, even once, had a drink of any type of alcoholic beverage? Please do not include times when you only had a sip or two from a drink.

1 Yes
2 No
DK/REF

3. HALINTRO

The next questions are about substances called hallucinogens. These drugs often cause people to see or experience things that are not real.

A list of some common hallucinogens is shown below. These and many other substances that people use as hallucinogens are often known by street names, and we can't list them all. Please take a moment to look at the substances listed below so you know what kind of drugs the next questions are about.

LSD, also called 'acid'
PCP, also called 'angel dust' or phencyclidine
Peyote
Mescaline
Psilocybin
'Ecstasy', also called MDMA
Ketamine, also called "Special K" or "Super K"
DMT, also called dimethyltryptamine
AMT, also called alpha-methyltryptamine
Foxy, also called 5-MeO-DIPT
Salvia divinorum

Press [ENTER] to continue

4. INHINTRO

These next questions are about liquids, sprays, and gases that people sniff or inhale to get high or to make them feel good.

We are not interested in times when you inhaled a substance accidentally — such as when painting, cleaning an oven, or filling a car with gasoline. The questions use the word 'inhalant' to include all the things listed below, as well as any other substances that people sniff or inhale for kicks or to get high.

Take a moment to look at the substances listed below so you know what kinds of liquids, sprays, and gases these questions are about.

Amyl nitrite, 'poppers,' locker room odorizers, or 'rush'
Correction fluid, degreaser, or cleaning fluid
Gasoline or lighter fluid
Glue, shoe polish, or toluene
Halothane, ether, or other anesthetics
Lacquer thinner, or other paint solvents
Lighter gases, such as butane or propane
Nitrous oxide or 'whippits'
Felt-tip pens, felt-tip markers, or magic markers
Spray paints
Computer keyboard cleaner, also known as air duster

Other aerosol sprays

Press [ENTER] to continue.

5. TR03

Please look at the names and pictures of the tranquilizers shown below.

PROGRAMMER: DISPLAY PILLS HERE FOR VALIUM, DIAZPEPAM, TRANXENE, AND OXAZEPAM.

In the **past 12 months**, which, if any, of these tranquilizers have you used?
To select more than one drug from the list, press the space bar between each number you have typed. When you have finished, press [ENTER].

1 Valium
2 Librium
3 Tranxene
4 Diazepam (generic)
5 Oxazepam (generic), also known as Serax
95 I have not used any of these tranquilizers in the past 12 months
DK/REF

6. TR05

Please look at the names and pictures of the tranquilizers shown below.

PROGRAMMER: DISPLAY PILLS FOR BUSPIRONE, HYDROXYZINE, AND MEPROBAMATE.

In the **past 12 months**, which, if any, of these tranquilizers have you used?
To select more than one drug from the list, press the space bar between each number you have typed.

When you have finished, press [ENTER].

1 Buspirone (generic), also known as BuSpar
2 Hydroxyzine (generic), also known as Atarax or Vistaril
3 Meprobamate (generic), also known as Equanil or Miltown
95 I have not used any of these tranquilizers in the past 12 months
DK/REF

7. ST04

Please look at the names and pictures of the stimulants shown below.

PROGRAMMER: DISPLAY PILLS FOR BENZPHETAMINE, DIDREX, DIETHYLPROPION, PHENDIMETRAZINE, AND PHENTERMINE.

In the **past 12 months**, which, if any, of these stimulants have you used?

To select more than one drug from the list, press the space bar between each number you have typed. When you have finished, press [ENTER].

1 Benzphetamine
2 Didrex
3 Diethylpropion
4 Phendimetrazine
5 Phentermine
95 I have not used any of these stimulants in the past 12 months
DK/REF

8. SV03

Please look at the names and pictures of the sedatives shown below.

PROGRAMMER: DISPLAY PILLS FOR DALMANE, HALCION, FLURAZEPAM AND TRIAZOLAM.

In the **past 12 months**, which, if any, of these sedatives have you used?

To select more than one drug from the list, press the space bar between each number you have typed. When you have finished, press [ENTER].

1 Dalmane
2 Halcion
3 Flurazepam (generic)
4 Triazolam (generic)
95 I have not used any of these sedatives in the past 12 months
DK/REF

9. PRINTROYR2

NOTE: For this question, we will assume a fill of 4 drug names.

[IF PR12MON=1 AND (PR11 NE 1 OR (PR11=1 AND PRYRCOUNT > 1))]

Earlier, the computer recorded that, in the **past 12 months**, you used [PRFILL].

Press Enter to continue.

PROGRAMMER: SHOW CALENDAR WITH 12-MONTH REFERENCE DATE FOR THE INTRO SCREEN

10. HLTH25

Below is a list of health conditions that you may have had during your lifetime.

Please read the list and type in the numbers of all of the conditions that a doctor or other health care professional has ever told you that you had.

To select more than one condition, press the space bar between each number you type. When you have finished, press [ENTER].

1 Any kind of heart condition or heart disease
2 Diabetes or sugar diabetes
3 Chronic bronchitis, emphysema, chronic obstructive pulmonary disease, also called COPD
4 Cirrhosis of the liver
5 Hepatitis B or C
6 Kidney disease, not including bladder infection or incontinence
7 Asthma
8 HIV or AIDS
9 Cancer or a malignancy of any kind
10 Hypertension, also called high blood pressure
95 None of the above - I have never had any of these conditions
DK/REF

11. AD19

[IF AD16 = 2, 3, 4, OR DK/REF] Once again, please think of times lasting two weeks or longer when [NUMPROBS] with your mood [WASWERE] most **severe and frequent**.

How often, during those times, was your emotional distress so severe that you **could not carry out your daily activities**?

1 Often
2 Sometimes
3 Rarely
4 Never
DK/REF

12. INTROINC

NOTE: For this question, we will assume a 4 person family with father, mother, son, daughter

[IF NO FAMILY MEMBERS IN ROSTER]

These next questions are about the kinds and amounts of income that you receive.

[IF ONE FAMILY MEMBER IN ROSTER AND HASJOIN NE 1]

These next questions are about the kinds and amounts of income received by you and your

[FAMILY RELATIONSHIP FILL].

[IF ONE FAMILY MEMBER IN ROSTER AND HASJOIN=1]

These next questions are about the kinds and amounts of income received by [SAMPLE MEMBER] and you.

[IF AT LEAST TWO FAMILY MEMBERS IN ROSTER AND HASJOIN NE 1]

These next questions are about the kinds and amounts of income received by your family living here, including you, your [FAMILY RELATIONSHIP FILLS].

[IF AT LEAST TWO FAMILY MEMBERS IN ROSTER AND HASJOIN=1]

These next questions are about the kinds and amounts of income received by [SAMPLE MEMBER] and [IF QD01=5 FILL his, QD01 = 9 FILL her] family living here, including you,

[IF QD01=5 FILL his, QD01 = 9 FILL her] [FAMILY RELATIONSHIP FILLS]. [PROGRAMMER NOTE: THE PROXY SHOULD NOT APPEAR IN [FAMILY RELATIONSHIP FILLS]. ALSO, USE 'other' AS A MODIFIER TO THE FAMILY RELATIONSHIP FILL WHEN THE RELATIONSHIP TYPE IS EQUAL TO PROXY RELATIONSHIP TYPE AND ONE OF THESE RELATIONSHIP TYPES IS STILL IN THE LIST. PLEASE PRECEDE EACH RELATIONSHIP WITH 'HIS/HER'.]

[IF HASJOIN NE 1] These questions refer to the calendar year [CURRENT YEAR - 1] rather than to the past 12 months that were referred to in some earlier questions. The calendar year [CURRENT YEAR - 1] would be from January 1st, [CURRENT YEAR - 1], through December 31st, [CURRENT YEAR - 1].

Press [ENTER] to continue.

A.3. Spanish Versions

1. CARD3a Tipos de Bebidas Alcohólicas

Cerveza
Cerveza
Cerveza ligera o 'lite'
Cerveza con poco alcohol (LA)
Malta con alcohol
'Ale'
Cerveza negra
Lager o Cerveza dorada

Vino
Vino tinto, blanco, rosado
'Wine coolers'
Champaña
Jerez
Vinos caseros, tales como uva moscatel, scuppernong o vinos frutales
Vinos fortificados tales como Cisco

Licor
Whisky americano
Ginebra o 'Gin'
Ron
Whisky escocés
Tequila
Vodka
Alcohol casero destilado ('Moonshine')

Licores de esencias, Cordiales y Brandy
Brandy
Cassis
Coñac
Crema de menta
Drambuie
Grand Marnier
Kahlua
Oporto
Schnapps
Tía María
Triple seco
Vermut

Bebidas mezcladas y Cocteles
Bloody Mary
Whisky con agua
Daiquiri
'Gin y tónic'
Manhattan
Margarita
Martini
Piña colada
Rob Roy
Ron con Coca Cola
Whisky escocés con soda
Whisky sour

Presione [ENTER] para continuar.

2. AL01

¿**Alguna vez** ha tomado una bebida alcohólica, aunque haya sido solo una vez? Por favor no incluya las ocasiones en que usted haya tomado solo uno o dos sorbos de una bebida.

1 Sí
2 No
DK/REF

3. HALINTRO

Las siguientes preguntas se tratan de las sustancias que se llaman alucinógenos. Estas drogas muchas veces hacen que las personas vean o experimenten cosas que no son reales.

A continuación hay una lista de algunos alucinógenos populares. Estas y muchas otras sustancias que la gente usa como alucinógenos se conocen frecuentemente por su nombre popular o de la calle. No podemos enumerarlos todos. Por favor preste atención al leer la lista de sustancias que sigue para saber a qué drogas se refieren las próximas preguntas.

LSD, también llamado 'ácido'
PCP, también llamado 'polvo de ángel' o fenciclidina
Peyote
Mescalina
Psilocibina
'Éxtasis,' también llamado MDMA
Ketamina, conocida en inglés como "Special K" o "Super K" y en español se le llama "Ketalar," "Hoyo K" o "vitamina K"
DMT, también llamado dimetiltriptamina
AMT, también llamado alfa-metiltriptamina
Foxy, también llamado "metoxi foxy" y cuyo nombre químico es 5-metoxi-N o 5-MeO-DIPT
Salvia divinorum, también llamada "Salvia de los adivinadores," "San Pedro," "planta sagrada" o "hierba pastora"

Presione [ENTER] para continuar.

4. INHINTRO

Las siguientes preguntas son acerca de líquidos, aerosoles o esprays y gases que las personas aspiran o inhalan para drogarse o para sentirse alegres.

No estamos interesados en ocasiones en que usted inhaló alguna sustancia accidentalmente como en el caso de pintar, limpiar un horno o echarle gasolina al automóvil. Las preguntas usan el término 'inhalante' para incluir todas las cosas mencionadas a continuación, así como cualquier otra sustancia que las personas aspiran o inhalan para divertirse o para drogarse. Por favor mire con atención la lista de sustancias a continuación, para saber a qué clases de líquidos, aerosoles o esprays y gases se refieren las próximas preguntas.

Nitrato de amilo, 'bombitas,' desodorante ambiental, o 'rush'
Líquido de corrección o 'liquid paper', desengrasador o líquido de limpieza
Gasolina o líquido para encendedores
Pegamento, crema o betún para limpiar zapatos, o tolueno
Halotano, éter u otros anestésicos
'Tiner' u otros solventes para pintura
Gases para encendedores, tales como butano o propano
Óxido nitroso o 'whippits'
Marcadores de punta fina, plumones o plumones mágicos

Pintura en aerosol
Limpiador para teclado de computadora, también llamado aire comprimido removedor de polvo
Otros aerosoles o esprays

Presione [ENTER] para continuar.

5. TR03

Por favor mire los nombres y las fotos de los tranquilizantes que se muestran a continuación.

PROGRAMMER: DISPLAY PILLS HERE FOR VALIUM, DIAZPEPAM, TRANXENE, AND OXAZEPAM.

En los **últimos 12 meses**, ¿cuál de estos tranquilizantes ha usado, si es que ha usado alguno?

Para seleccionar más de un medicamento en la lista, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

1 Valium
2 Librium
3 Tranxene
4 Diazepam (genérico)
5 Oxazepam (genérico), también conocido como Serax
95 No he usado ninguno de estos tranquilizantes en los últimos 12 meses
DK/REF

6. TR05

Por favor mire los nombres y las fotos de los tranquilizantes que se muestran a continuación.

PROGRAMMER: DISPLAY PILLS HERE FOR BUSPIRONE, HYDROXYZINE, AND MEPROBAMATE.

En los **últimos 12 meses**, ¿cuál de estos tranquilizantes ha usado, si es que ha usado alguno?

Para seleccionar más de un medicamento en la lista, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

1 Buspirona (genérico), también conocido como BuSpar
2 Hidroxizina (genérico), también conocido como Atarax o Vistaril
3 Meprobamato, (genérico) también conocido como Equanil o Miltown
95 No he usado ninguno de estos tranquilizantes en los últimos 12 meses
DK/REF

7. ST04

Por favor mire los nombres y las fotos de los estimulantes que se muestran a continuación.

PROGRAMMER: DISPLAY PILLS FOR BENZPHETAMINE, DIDREX, DIETHYLPROPION, PHENDIMETRAZINE, AND PHENTERMINE.

En los **últimos 12 meses**, ¿cuál de estos estimulantes ha usado, si es que ha usado alguno?

Para seleccionar más de un medicamento en la lista, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

1 Benzfetamina
2 Didrex
3 Dietilpropión
4 Fendimetracina
5 Fentermina
95 No he usado ninguno de estos estimulantes en los últimos 12 meses
DK/REF

8. SV03

Por favor mire los nombres y las fotos de los sedantes que se muestran a continuación.

PROGRAMMER: DISPLAY PILLS FOR DALMANE, HALCION, FLURAZEPAM AND TRIAZOLAM.

En los **últimos 12 meses**, ¿cuál de estos sedantes ha usado, si es que ha usado alguno?
Para seleccionar más de un medicamento en la lista, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

1 Dalmane
2 Halcion
3 Flurazepam (genérico)
4 Triazolam (genérico)
95 No he usado ninguno de estos sedantes en los últimos 12 meses
DK/REF

9. PRINTROYR2

NOTE: For this question, we will assume a fill of 4 drug names

[IF PR12MON=1 AND (PR11 NE 1 OR (PR11=1 AND PRYRCOUNT > 1))]

Anteriormente, la computadora registró que usted usó [PRFILL] en los últimos 12 meses.

Presione [Enter] para continuar.

PROGRAMMER: SHOW CALENDAR WITH 12-MONTH REFERENCE DATE FOR THE INTRO SCREEN

10. HLTH25

A continuación se muestra una lista de trastornos de la salud que usted pudiera haber tenido en el **transcurso de su vida.**

Por favor, lea la lista y escriba los números correspondientes a las enfermedades que **alguna vez** un doctor u otro profesional médico le dijo que tuvo.

Para seleccionar más de una enfermedad, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

1 Algún tipo de enfermedad o trastorno del corazón
2 Diabetes o diabetes del azúcar
3 Bronquitis crónica, enfisema, enfermedad pulmonar obstructiva crónica, también llamada COPD en inglés
4 Cirrosis del hígado
5 Hepatitis B o C
6 Enfermedad de los riñones, sin incluir infección a la vejiga o incontinencia urinaria
7 Asma
8 VIH o SIDA
9 Cáncer o algún tipo de tumor maligno
10 Hipertensión, también llamada presión sanguínea alta
95 Ninguna enfermedad arriba mencionada - Nunca he tenido ninguna de estas enfermedades
DK/REF

11. AD19

[IF AD16 = 2, 3, 4, OR DK/REF] Una vez más, por favor piense en las veces que [NUMPROBS] con su estado de ánimo [WASWERE] por dos semanas o más.

Durante esas ocasiones, ¿con qué frecuencia era su malestar emocional tan grave que **no podía realizar sus actividades diarias**?

1 Muchas veces
2 Algunas veces
3 Casi nunca
4 Nunca
DK/REF

12. INTROINC

NOTE: For this question, we will assume a 4 person family with father, mother, son, daughter

[IF NO FAMILY MEMBERS IN ROSTER]

Las siguientes preguntas se tratan de los tipos de ingreso y las cantidades que usted recibe.

[IF ONE FAMILY MEMBER IN ROSTER AND HASJOIN NE 1]

Las siguientes preguntas se tratan de los tipos de ingreso y las cantidades que reciben usted y su [FAMILY RELATIONSHIP FILL].

[IF ONE FAMILY MEMBER IN ROSTER AND HASJOIN=1]

Las siguientes preguntas se tratan de los tipos de ingreso y las cantidades que reciben [SAMPLE MEMBER] y usted.

[IF AT LEAST TWO FAMILY MEMBERS IN ROSTER AND HASJOIN NE 1]

Las siguientes preguntas se tratan de los tipos de ingreso y las cantidades que reciben los miembros de su familia que viven aquí, incluyéndose usted, su [FAMILY RELATIONSHIP FILL].

[IF AT LEAST TWO FAMILY MEMBERS IN ROSTER AND HASJOIN=1]

Las siguientes preguntas se tratan de los tipos de ingreso y las cantidades que reciben [SAMPLE MEMBER] y los miembros de su familia que viven aquí, incluyéndose usted, [FAMILY RELATIONSHIP FILL] de su [SAMPLE MEMBER].

[PROGRAMMER NOTE: THE PROXY SHOULD NOT APPEAR IN [FAMILY RELATIONSHIP FILLS]. ALSO, USE 'otro' AS A MODIFIER TO THE FAMILY RELATIONSHIP FILL WHEN THE RELATIONSHIP TYPE IS EQUAL TO PROXY RELATIONSHIP TYPE AND ONE OF THESE RELATIONSHIP TYPES IS STILL IN THE LIST. PLEASE PRECEDE EACH RELATIONSHIP WITH 'SU'.]

[IF HASJOIN NE 1]Estas preguntas se refieren al año calendario [CURRENT YEAR-1] en vez de los últimos 12 meses a los que se refirieron algunas preguntas anteriores. El año calendario [CURRENT YEAR-1] es del 1 de enero de [CURRENT YEAR-1] hasta el 31 de diciembre del año [CURRENT YEAR-1].

Presione [ENTER] para continuar.