2020

A Comparison of Computer- and Interviewer-Administered Measures to Identify Heavy/Problem Alcohol and Other Drug Use in Primary Care Patients

David J. Pomm
Virginia Commonwealth University

A COMPARISON OF COMPUTER- AND INTERVIEWER-ADMINISTERED MEASURES TO IDENTIFY HEAVY/PROBLEM ALCOHOL AND OTHER DRUG USE IN PRIMARY CARE PATIENTS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

by

DAVID JONATHAN POMM, M.S.
Master of Science, Clinical Psychology
Virginia Commonwealth University, 2016

Director: Dace S. Svikis, Ph.D.
Professor, Department of Psychology
Deputy Director, VCU Institute for Women's Health

Virginia Commonwealth University
Richmond, VA
November 20, 2020

Acknowledgments

The completion of this dissertation requires an enthusiastic expression of appreciation to my graduate mentor and dissertation chair Dace Svikis. I am very grateful for Dr. Svikis's continuous support and guidance during my graduate training and on this project. The numerous invaluable experiences and discussions with Dr. Svikis have helped me grow exponentially as a researcher. I would also like to thank the members of my dissertation committee, Bruce Rybarczyk, Paul Perrin, Pamela Dillon, and Arline Bohannon, for their guidance and feedback on this project. This research is stronger because of your expertise and enthusiastic support. In addition, I would like to thank Dr. Leroy Thacker for providing statistical consultation throughout the design and completion of this study. I would like to express deepest gratitude to my family for their continuous support throughout my academic endeavors. Most importantly, I must thank my wife, Brianna, for her patience and many sacrifices during this journey. Thank you for sharing the first years of our marriage with VCU. I could not have accomplished this chapter without your inspiration, encouragement, and steadfast support.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

Abstract

A COMPARISON OF COMPUTER- AND INTERVIEWER-ADMINISTERED MEASURES TO IDENTIFY HEAVY/PROBLEM ALCOHOL AND OTHER DRUG USE IN PRIMARY CARE PATIENTS

By David J. Pomm, M.S.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2020

Major Director: Dace S. Svikis, Ph.D.
Professor, Department of Psychology
Deputy Director, VCU Institute for Women's Health

Substance use is a leading cause of morbidity and mortality that is under-identified in primary care settings. Screening for substance use in primary care requires an efficient and accurate approach. Exploring methods to collect sensitive and accurate data about substance use and related problems is important to inform research practice and clinical care. Using RCT data comparing computerized and therapist delivered brief intervention for substance use problems, the present study had a unique opportunity to compare computerized anonymous and confidential surveys to a confidential, interviewer-administered assessment in a sample of N = 540 males and females at risk for heavy/problem alcohol and/or drug use and recruited through an urban primary care clinic. This study also compared self-reports of recent substance use to urine drug assay findings. Finally, the study identified correlates of any recent (past 30-day) use and days of alcohol and other drug use per week in the past 30 days. The sample was 39% male, 78% African American, and had a mean age of 45.1 years. More participants self-reported alcohol use on the computerized, anonymous screen, including any recent (past 30-day) alcohol use, binge drinking, and problems associated with alcohol use. Any recent (past 30-day) illicit

drug use rates were highest on the confidential computerized survey, and quantity/frequency of alcohol use as well as frequency of illicit drug and prescription drug misuse were highest on the interviewer-administered assessment. Overall concordance rates between interviewer-administered assessment and urine drug screening (UDS) were 72% or higher for each substance, driven by large subgroups with no use. Among participants with discordant use, marijuana and heroin / opiate use were the only substance with lower detection on UDS than self-report. Exploratory analyses examined psychosocial correlates of self-reported substance use. Anonymously screening for recent substance use followed by an interviewer-administered assessment provides the most parsimonious method to identify sensitive data about substance use and related behaviors in primary care. This approach has the potential to facilitate implementation of substance screening into demanding clinical environments.

A Comparison of Computer- and Interviewer-Administered Measures to Identify Heavy/Problem Alcohol and Other Drug Use in Primary Care Patients

Substance misuse constitutes a major public health problem; one that costs the United States (U.S.) over $740 billion annually through health care costs, lost work productivity, and crime (NIDA, 2017). Substance misuse can result in a diagnosis of substance use disorder (SUD) or a person becoming addicted, necessitating evidence-based treatment services. However, of the estimated 21.2 million people age 12 and older (7.8%) in need of substance use treatment, many do not engage in requisite specialty treatment (Substance Abuse and Mental Health Services Administration [SAMHSA], 2019) and are under-recognized in medical settings (Edlund, Unutzer, & Wells, 2004; Rehm et al., 2016). In efforts to destigmatize and explain more fully the complexity of this chronic disease, the American Society of Addiction Medicine (ASAM) redefined addiction in 2019. The updated definition reads:

> Addiction is a treatable, chronic medical disease involving complex interactions among brain circuits, genetics, the environment, and an individual's life experiences. People with addiction use substances or engage in behaviors that become compulsive and often continue despite harmful consequences. Prevention efforts and treatment approaches for addiction are generally as successful as those for other chronic diseases. (p 2).

There is growing recognition that evidence-based treatment services can play important roles in the full spectrum of addiction care, with recent efforts focusing on the provision of brief, evidence-based nonpharmacologic therapy for SUDs, such as Screening, Brief Intervention, and Referral to Treatment (SBIRT) (Babor et al., 2017; Broyles et al., 2013; Jhanjee, 2014). Screening for substance use in adult primary care settings is recommended by the United States Preventive Services Task Force (USPSTF), ranking as the top 10 highest prevention priority for

adults in the U.S (USPSTF, 2018; USPSTF, 2019). The U.S. Surgeon General's report on addiction recommends screening for other drug use as well as alcohol, which is the current approach of federally funded SBIRT programs (HHS, 2016; SAMHSA, 2020a).

A key component of the SBIRT approach is the linking of screening and assessment results with appropriate early intervention services. The purpose of screening is to identify individuals whose levels, patterns, or consequences of substance use needs further assessment or referral to services based on risk level (SAMHSA, 2020a). Studies often use self-report to measure alcohol and drug use, but under-reporting is common. A growing body of literature has identified several factors that may affect the reliability and accuracy of self-reports, such as degree of anonymity, extent to which the behavior assessed is illicit or socially undesirable, the method by which data is collected (face-to-face interview or questionnaire), and whether biological verification is planned (Chermack et al., 2000; Bone et al., 2016; Fleming et al., 2007; Harrison & Hughes, 1997).

Further, commonly used methods for screening and assessing substance use are designed with different purposes and vary widely in the time frame they capture; thus, each has specific strengths and weaknesses (Kaner et al., 2009; O'Conner et al., 2018). This variation in screening practices complicates the ability to clearly define best practices. Structured interview-administered substance use screening and assessment approaches are thought to be most accurate, but they are challenging to implement in practice because they require staff time and training (Williams et al., 2011). Semi-structured, interviewer-administered instruments that use supplemental memory aids during the assessment process (e.g., Timeline-Follow-Back, TLFB) (Sobell & Sobell, 1992) to facilitate recall demonstrate excellent reliability and validity (Rosenbaum et al., 2006), but require staff time and training. Brief screening tools have been

developed to efficiently identify alcohol and drug use, but these do not provide enough information about the specific substances used, or the patient's risk level, to guide clinical actions. Biologic measures also have limitations, including high cost, varying windows for detection of different drugs, and ease of collection (ASAM, 2017).

Technology provides new possibilities for standalone or facilitated screening. For example, Audio Computer-Assisted Interview (ACASI) technology allows for self-administered method that facilitate confidential administration, allowing patients to respond with lower threat of social desirability bias (Dolezal et al., 2012; Estes et al., 2010; Richter & Johnson, 2001). However, the impact of ACASI technology and degree of privacy (e.g., anonymous vs. confidential) on the way respondents answer questions across alcohol and illicit drug use outcomes, including prescription drug misuse, has largely been unexplored. Also, little is known about the agreement between interviewer-administered screening and assessment questionnaires and other questions on the frequency of substance use administered via ACASI. The process of validating self-report against self-report leaves opportunities for respondents to deny substance use in both conditions (Ondersma et al., 2012); however, substantial disagreement would indicate that studies using one method should not be compared to similar studies that used the other method, impairing the field's ability to aggregate data across studies.

The present study compares self-reports of alcohol and illicit drug use, and prescription drug misuse across different screening measures and methods of administration and across varying levels of privacy. This study also compares self-reports of recent substance use to urine drug toxicology findings. Data were part of a randomized clinical trial (RCT) comparing computerized and therapist delivered brief intervention for substance use problems. Participants were identified through an urban primary care clinic using an anonymous, computer-

administered health screen (CAHS) with embedded questions about alcohol and other drug use. Data was collected on a diverse range of variables including patient demographics, drug and alcohol use, family history of substance use, living environment, and social supports. Participants (N=713) meeting criteria for heavy/problem substance use were randomized to one of four study arms: CAHS only (standard care true control), confidential, computerized assessment only intervention (CA), confidential, computerized assessment plus computer-delivered brief intervention (CACI), or confidential, computerized assessment plus therapist-delivered brief intervention (CATI).

The present study was conducted with the baseline data for 540 individuals who completed the CAHS and were randomized to three of the four study arms. Those in the standard care control group were excluded as they did not complete the measures examined in the present study. This research directly compares an anonymous screen, confidential, computerized assessment, and confidential, interviewer-administered assessment. Other psychosocial and health disparity variables (e.g., age, gender, race, marital status, years of education, ethnicity, number of medical conditions) were also examined to identify of self-disclosure of any alcohol and other drug use as well as substance use per week in the past 30-days.

The present study had four primary aims:

I. Aim 1: To compare rates of participant self-disclosure of alcohol and illicit drug use and prescription drug misuse obtained by an anonymous, computer-administered health screen to those obtained by a confidential, computerized assessment and interviewer-administered research assessment. Four hypotheses were tested:

a. Hypothesis 1: Participants will report more recent binge drinking days (past 30) on the anonymous, computer-administered screener (CAHS) compared to the confidential, interviewer-administered assessment (IARA).

b. Hypothesis 2: Participants will report consuming, on average, more drinks per week over the past 30 days by anonymous, computer-administered screener (CAHS) compared to the confidential, computerized research assessment (CARA). The study also hypothesized that participants will report consuming, on average, more drinks per week by CAHS compared to the interview-administered assessment (IARA).

c. Hypothesis 3: Participants will report, on average, more days of drug use per week over the past 30 days by anonymous, computer-administered screener (CAHS) compared to confidential, computerized assessment (CARA). The study also hypothesized that participants will report, on average, more days of drug use per week by CAHS compared to the interview-administered assessment (IARA).

d. Hypothesis 4: Participants will report, on average, more days of prescription drug misuse per week by anonymous, computer-administered screener (CAHS) compared to confidential, computerized assessment (CARA). The study also hypothesized that participants will report, on average, more days of drug use per week by CAHS compared to the interview-administered assessment (IARA).

II. Aim 2: Compare endorsement of alcohol and/or drug use related problems between the anonymous, computer-administered screener and confidential, computerized assessment.

a. Hypothesis 1: Participants will be more likely to score positive for problematic substance use on the anonymous, computer-administered screener compared to the confidential, computerized assessment.

III. Aim 3: To compare agreement between self-report of recent drug use and biological measures (i.e., urinalysis).

a. Hypothesis 1: Proportion of illicit drug use and non-medical prescription drug use will be higher by urine assay compared to self-report.

IV. Aim 4: To identify correlates (e.g., age, gender, race, marital status, years of education, ethnicity, number of medical conditions, mental health-related conditions, etc.) of self-disclosure of any alcohol and other drug use as well as substance use per week in the past 30-days.

Review of the Literature

Epidemiology of Substance Use Disorders

Substance misuse, defined as the use of alcohol or drugs in a manner, situation, amount, or frequency that could cause harm to the user or those around them (The U.S. Department of Health and Human Services [HHS], 2016) is a problem that affects millions of Americans (Grant et al., 2017; Rudd et al., 2016; Stahre et al., 2014). The economic, psychosocial and public health impact of increased mortality and morbidity is substantive (Esser et al., 2020; Gomes et al., 2018; HHS, 2016; Johnson et al., 2014; Peacock et al., 2018; Rehm, 2020; Smith et al., 2015).

The National Survey on Drug Use and Health (NSDUH) in 2018, the most recent year for which drug use data are available, found that 16.6 million Americans reported heavy drinking in the past month and 31.9 million people were users of at least one illicit drug (SAMHSA, 2019). Also, in 2018 an estimated 16.9 million Americans misused prescription drugs at least once in

the past year and nearly 20.3 million met diagnostic criteria for a substance use disorder (SUD) (alcohol or illicit drugs) (SAMHSA, 2019).

Furthermore, binge-level drinking (defined as consuming five or more drinks on an occasion for men and four or more drinks on an occasion for women) contributes largely to alcohol-induced mortality among U.S. adults (Kanny et al., 2020; National Institute on Alcohol Abuse and Alcoholism [NIAAA], 2017), and significant increases in the prevalence of binge drinking have been observed in recent years (Grucza et al., 2018) with 1 in 4 people (67.1 million) binge drinking in 2018 (SAMHSA, 2019).

The U.S. is also currently facing an opioid epidemic. In 2018, there was an estimated 2.0 million people who had an opioid use disorder (SAMHSA, 2019). Also, in 2018, the Centers for Disease Control (CDC) estimated that 70% of drug overdose deaths involved an opioid. Synthetic opioids are of particular concern with overdose deaths from fentanyl increasing 10% from 2017 to 2018 (Scholl et al., 2019; Wilson et al., 2020).

Receipt of Services for Substance Use

There are myriad of nonpharmacological evidence-based interventions for the treatment of SUDs, ranging from individual and group psychotherapies, to self-control and social skills training, to aversion therapies (Hester & Squires, 2004). Please see Polak et al. (2020) for a comprehensive review of evidence-based behavioral and psychosocial interventions for the treatment of alcohol and other SUDs. Unfortunately, of the approximately 18.9 million people in need for substance use treatment in 2018, only 20% utilized any SUD treatment services, and only 13% received specialty treatment (SAMHSA, 2019). The vast majority (94.9%) of those who needed substance use treatment but did not receive specialty treatment did not think they needed treatment (SAMHSA, 2019). In general, reasons for not receiving substance use

treatment include not being ready to stop using, lack of health care coverage (SAMHSA, 2019), believing they don't have a problem (Edlund et al., 2006; Edlund et al., 2009; Park-Lee et al., 2017) or reporting reservations about seeking treatment (Cohen et al., 2006; Grant, 1997; Kim et al., 2017; Perron et al., 2009).

Screening, Brief Intervention, and Referral to Treatment (SBIRT)

In 2003, SAMHSA launched a major initiative—SBIRT— a public health model designed to address substance misuse of all individuals presenting for care but not specifically seeking treatment for substance use (Madras et al., 2009).

The major elements of SBIRT are 1) screening (identifying problematic substance use with standardized measures), 2) brief intervention for people who have problematic or hazardous substance use problems (provider and patient engage in evidenced-based counseling to raise patient's awareness of substance use and its consequences and to elicit positive behavior change), and 3) referral to treatment for those with more severe substance use problems (Agerwala & McCance-Katz, 2012; Babor et al., 2007; Babor & Higgins-Biddle, 2001; Polak et al., 2020; SAMHSA, 2020a).

SBIRT can be applied in a range of clinical settings, ranging from primary care clinics to hospital emergency departments, and varied community-based care centers. Together, this health network provides access to patient populations generally not seen in traditional SUD treatment facilities. Healthcare providers can implement SBIRT during routine medical visits by asking patients about their substance use and delivering brief intervention to those who screen at risk, with referral to treatment when warranted (Babor et al., 2007; Hargraves et al., 2017).

Effectiveness of Alcohol SBIRT in Primary Care

A substantial number of systematic reviews and meta-analyses have demonstrated the effectiveness of SBIRT in primary care settings for people drinking in excess to established safety guidelines (Álvarez-Bueno et al., 2015; Cuijpers et la., 2004; Hettema et al., 2005; Jonas et al., 2012; Kaner et al., 2018; Moyer et al., 2002; Solberg et al., 2008). The effectiveness of SBIRT for those with more severe alcohol problems or alcohol dependence, however, is less clear (McCambridge & Satiz, 2017; Saitz, 2010; Saitz, 2013).

Taken together, the findings summarized support moderate beneficial effect of screening and brief intervention for unhealthy alcohol use in adults. This evidence led the USPSTF to recommend screening and brief interventions to reduce unhealthy alcohol use by adults in primary care (Moyer, 2013; Reus et al., 2018; USPSTF, 2018). The USPSTF also recommends screening for unhealthy drug use in adults age 18 years or older, noting that screening should be implemented when effective treatment, including brief interventions, can be offered or referred (USPSTF, 2019; USPSTF, 2020). This recommendation statement replaces the 2008 USPSTF recommendation which concluded that the evidence at that time was insufficient to recommend screening and treatment for other drugs in adults.

Effectiveness of Drug SBIRT in Primary Care

Adapting SBIRT to target other drugs of abuse has been challenging. For screening, there is a need to target a range of drugs, including prescription drugs, not simply one substance. Further, the health consequences are less well-defined than for alcohol, and differ by class of drug. Finally, illicit drug use carries greater social stigma and greater fear of negative sanctions or consequences compared to heavy alcohol use. SBIRT outcomes, therefore, may be different

for other drugs compared to those targeting heavy and problem drinking (McCambridge & Rollnick, 2014; Saitz, 2014).

SBIRT for other drugs of abuse has yielded mixed results, with some studies finding screening and brief intervention to be effective, with patient reductions in illicit drug use (Bernstein et al., 2005; Humeniuk et al., 2008; Humeniuk et al., 2012; Krupski et al., 2012; Madras et al., 2009). Others, however, have argued that negative findings predominate, with several RCTs finding the progression of brief intervention from efficacy to effectiveness for drugs other than alcohol has exceeded its evidence base (Bogenschutz et al., 2014; Hingson & Compton, 2014; Saitz et al., 2010; Saitz et al., 2014a; Roy-Byrne et al., 2014; Young et al., 2014). Based on the current literature, dissemination of drug SBIRT has been limited, however, identifying and addressing illicit drug use and prescription drug misuse in primary care patients remains essential (Pace & Uebelacker, 2018; Tong et al., 2019).

Estimate of Potential Cost Benefit

Substance abuse is a serious disease that adversely affects U.S. economic vitality. It is estimated that alcohol, prescription opioids, and illicit drug abuse cost the U.S. $740 billion annually (Kasunic & Lee, 2014; NIDA, 2017). The cost of substance abuse treatment is approximately $11,600 per user, resulting in an economic burden of $224 billion, not including $84 billion lost in overall healthcare produced by the adverse outcomes of substance abuse (Florence et al., 2018; Sacks et al., 2015).

Data suggest that primary care patients with SUDs often have multiple comorbidities associated with high clinical complexity and cost (Falk et al., 2008; Gerteis et al., 2014; Hayes et al., 2016; Korthuis et al., 2017; Saitz & Daaleman, 2017; Skinner et al., 2016). Patients with

chronic conditions and a co-occurring SUD are complicated to treat in primary care (Center for Behavioral Health Statistics and Quality, 2016; Graham et al., 2017).

To promote SUD related care in primary care, SUDs have been re-conceptualized as chronic medical conditions in need of evidence-based screening and treatment (ASAM, 2019; Shapiro et al., 2013). SBIRT programs were included in the essential health benefits package as a part of the Affordable Care Act passed in 2010 (Moyer, 2013), and continuing to implement these evidence-based programs to identify at-risk users may mitigate the economic cost of specialized service provision. Economically, screening and brief interventions (SBI) have considerably lowered costs compared to conventional SUD treatment (Barbosa et al., 2015; Bray et al., 2014; Zarkin et al., 2015). A review by Bray et al., (2012) found costs for SBI for alcohol ranged from $0.51 to $601.50 per screen and from $3.41 to $243.01 per brief intervention. Other economic analyses on per-patient screening costs for alcohol and drug use is relatively limited, and data are needed with a focus on improving efficiency, such as computerized screening (Cowell et al., 2017).

Therefore, with the support from Triple-Aim reforms and policy shifts toward value-based care, allocating resources to investigate the magnitude of SUDs and associated comorbidities, and implementing screening and care-coordinating approaches for high-need, high-risk patients will help to improve health outcomes and reduce overall healthcare costs.

Substance Use Screening and Assessment Measures

Reliable and valid measures to identify individuals with heavy/problem substance use are an essential part of the public health armamentarium in the prevention and treatment of substance use disorders. Many standardized measures focus on quantity and frequency (QF) of use, with others targeting problems and consequences associated with use. Methods of administration also

vary, ranging from pencil-and-paper self-administered surveys and face-to-face interviews (IA) to computer/online self-administered and Audio Computer-Assisted Interviews (ACASI). The latter provides a user-friendly computer interface that guides people through a survey, using digitally recorded instructions, questions, and answers (Lavrakas, 2008). Variability also exists between alcohol and drug use screening and assessment. For example, reliance on self-report data via a broad array of measures is typical (Babor et al., 2000) but concurrent biological testing may also be used (Cone, 1997; Donovan et al., 2012). The window of assessment can vary from recent (typically past week or 30 days), to past 3 months, past year, and lifetime (any) use. The intention of these instruments' ranges from assessment of general use, screening for problems, and facilitating diagnostic assessment. Intended use of a measure also plays a role. In particular, primary care settings require tools that are efficient and accurate to identify individuals at risk who would not necessarily be seeking treatment for a substance use disorder (Saunders et al., 1993; SAMHSA, 2020a). However, other settings may require measures that provide detailed information about the specific substances used and the patient's risk level to guide clinical actions or inform research effectiveness.

In general, self-report remains a key metric for substance use screening, particularly in drug treatment programs and health care practices (Connors & Volk, 2004; Del Boca & Darkes, 2003; Richter & Johnson, 2001; Sobell & Sobell, 2004), but also often in clinical research (Clark et al., 2016). Such measures are inexpensive, easy to administer, and widely accessible, especially when compared to biological measures such as urine or oral fluid testing. Self-report measures have also consistently demonstrated good accuracy in research studies. Table 1 summarizes key characteristics of measures used in primary care settings looking separately at alcohol, other drugs, and alcohol and other drugs (see Babor & Kadden (2005) for a review).

Table 1. *Commonly Used Substance Use Screening and Assessment Measures in Primary Care Settings.*

Test, Definition, and Citation	Scoring	Pros	Cons
Alcohol			
Alcohol Use Disorders Identification Test (AUDIT) (Babor et al., 2001)	-10 questions; scoring: (> 8) indicate alcohol abuse/ dependence	-Identifies hazardous drinking & dependence -Lifetime/past year use	-Sensitivity low in elderly populations
Alcohol Use Disorders Identification Test – C (AUDIT-C) (Bush et al., 1998)	-3 questions; scoring: positive scores equal > 3 for women and > 4 for men	-Brief -Focuses on quantity, frequency, and pattern of drinking	-Different scoring for men/women -Less sensitive and specific for recent use
CAGE (Ewing, 1984) questionnaire; acronym for: **C**ut Down, **A**nnoyed, **G**uilty, **E**ye-opener	-4 questions; scoring: Indication of alcohol problems (≥2)	-Brief -Identifies potential alcohol problems	-Lifetime assessment only; potential for falsification
Michigan Alcoholism Screening Test (MAST) (Selzer, 1971)	-24 yes/no questions: Probable alcoholism (≥ 5)	-Identifies potential alcohol problems and dependence	-Lengthy -Absence of early identification -Potential for falsification
T-ACE (Sokol et al., 1989) questionnaire; acronym for: **T**olerance**, A**nnoyance, **C**ut Down, **E**ye-opener	-4 questions; scoring: Indication of alcohol problems (≥2)	-Brief -Screens for risky drinking during pregnancy -Lifetime/current use	-For women only
Drugs			
Screen of Drug Use (SoDU) (Tiet et al., 2015)	-2 questions; scoring: positive if ≥ 7 days of drug use in past 12 months and 2nd question is omitted (i.e., asking using drugs more than they meant to)	-Easy interpretation -Days of drug use in the past 12 months -Identifies negative consequences of drug use	-Direct obvious or face valid questions -Validated in Veteran's Affairs population only

Drugs & Alcohol			
Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) (WHO, 2002); semi-structured interview	8-57 questions; risk score for each substance (low-moderate- or high)	-Identifies risky alcohol/drug use and alcohol/drug problems -Lifetime & past 90-day use -Strong diagnostic accuracy -Self-administered computerized version	-Lengthy; structured interview -Complex scoring -Skip patterns
CAGE-AID (Brown & Rounds, 1995) questionnaire, expanded CAGE to include alcohol and other drugs	-4 questions; scoring: Probable alcoholism (> 2)	-Brief -Identifies potential alcohol/drug problems	- Lifetime assessment only; potential for falsification -More sensitive but less specific than the CAGE
The Tobacco, Alcohol, Prescription medication, and other Substance use tool (TAPS) (McNeely et al., 2016)	Two components: TAPS-1 – 4-item screen; any response other than 'never' = positive. TAPS-2 – brief assessment; risk score for each substance (No use, problem use, higher risk)	-Combines screening (frequency of use in the past 12-months) and brief assessment (past 3-month problem use) -Self-administered online or interviewer-administered -Provides substance-specific risk information on current use	-Automatic generated risk levels for each substance class. -Less sensitive for prescription medications -Final score based on past 3-month use/problems. -English version only
Timeline-Follow-Back (TLFB) (Sobell & Sobell, 1992)	N/A	-Focuses on quantity, frequency, and pattern of drinking/drug use -Estimates past 7 days to 2 years of use -Use of supplemental memory aids	-Lengthy -Retrospective estimates of use -Difficult to identify problems or dependence -Interview required

Alcohol Screening

For alcohol, many instruments with good psychometric properties are available for the identification of at-risk or harmful drinking and associated problems (Connors & Volk, 2004;

O'Connor et al., 2018; Saitz, 2007; Sobel & Sobel, 2004). The better-known instruments are summarized below, highlighting their psychometric properties as well as their scientific and practical consideration for their use.

Alcohol Use Disorders Identification Test (AUDIT)

The 10-question AUDIT, developed by the World Health Organization (WHO), is widely recommended and is often considered the "gold standard" among alcohol screening measures (Babor et al., 2001). The AUDIT focuses on both at-risk drinking (current to past year) as well as alcohol-related consequences and problems. Questions also cover identification of heavy drinking (drinking more than 6 standard drinks on any one occasion) and common symptoms of alcohol dependence. The AUDIT's psychometric properties are well established. Internal reliability has been consistently strong, with Chronbach's alpha scores in the range of 0.80-0.94. At the recommended cut-off of 8, it is most effective in identifying subjects with at-risk, hazardous, or harmful drinking (sensitivity, 51%-97%; specificity, 78%-96%). Other studies report high internal consistency and reliability (r=.86) and good sensitivity but somewhat lower, specificity, for ICD-10 alcohol use disorders (Babor et al., 2001; Berner et al., 2007; de Meneses-Gaya et al., 2009).

Disadvantages of the AUDIT include length (10-items), with 2+ minute administration time and more complex scoring than other screeners. To address this, an abbreviated version, the Alcohol Use Disorders Identification Test – C (AUDIT-C, Bush et al., 1998) with 3 items was developed and tested. Focusing only on consumption and patterns of drinking (sensitivity 0.86, specificity 0.89 in men; sensitivity 0.73, specificity 0.91 in women) the AUDIT-C is particularly useful in screening for risky/hazardous drinking (Dawson et al., 2012). The USPSTF (2018)

considers the AUDIT-C as another instrument of choice for screening for unhealthy alcohol use in primary care settings.

CAGE

Developed in the late 1960s, the four-item CAGE (Ewing, 1984) screens for alcohol-related problems. CAGE is a mnemonic that stands for Cut down, Annoyed, Guilty, and Eyeopener. The four CAGE questions are: 1) Have you ever felt you should Cut down on your drinking?, 2) Have people Annoyed you by criticizing your drinking?, 3) Have you ever felt bad or Guilty about your drinking?, and 4) Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover (Eye-opener)? Each affirmative response receives one point with total scores ranging from zero to four. Typically, a score of ≥ 2 points on the CAGE is used as the cut-off for risk of problem drinking (Ewing, 1984). The CAGE is commonly used in primary care settings (Tan et al., 2018). The screener has good validity to screen for alcohol dependence in primary care patients (sensitivity, 43%-94%; specificity, 70%-97%) (Fiellin et al., 2000). It also has demonstrated high test-retest reliability (0.80-0.95) (Dhalla & Kopec, 2007).

A common criticism of the CAGE is that it is less sensitive for identifying non-dependent at-risk drinkers. Gender and ethnicity have also been found to affect its performance, with studies showing a sensitivity as low as 50% in adult white women and as low as 39% in at-risk groups ages 60 and over (Connors & Volk, 2004). The questionnaire should not precede questions on alcohol quantity or frequency as its sensitivity is dramatically improved by an open-ended introduction (Steinweg & Worth, 1993). Finally, the CAGE asks about "lifetime" experience rather than current drinking, so a person who no longer drinks may screen positive unless the clinician directs the questions to focus on a more current timeframe; therefore, follow

up with the AUDIT of AUDIT-C is recommended for those who score positive on the CAGE (USPSTF, 2018).

Michigan Alcoholism Screening Test (MAST)

One of the first alcohol screening measures, the MAST (Selzer, 1971), consists of 24 yes-no questions focused on the medical, social, family and legal consequences of alcohol use. Items are weighted (1 to 5 points) and summed, and reliability and validity have been established with cutoff scores for probable (> 4 points) and definite (> 5 points) alcoholism. The MAST provides a gross, general measure of lifetime problem drinking severity and has been used to guide choice of treatment modality as well as further inquiry into alcohol-related problems (Zung, 1982).

The MAST has been criticized because of its length, risk for underreporting or falsification, failure to discriminate between past and present drinking, and a focus on finding cases of alcohol dependence rather than early problem identification. Also, certain items on the Michigan Alcoholism Screening Test (MAST) fail to specify referent or provide drinking norms, which can confuse interpretation (Svikis et al., 1991). To address the issues of length, versions have been developed, including the 10-item Brief MAST (bMAST) (Pokorny et al., 1972) and the 13-item Short MAST (SMAST) (Selzer et al., 1975). Both increase feasibility of screening in clinical settings but still maintain a focus on identifying active dependence.

T-ACE

While alcohol use prevalence and comorbid conditions differ in men and women, efforts to develop sex/gender-specific screening instruments have focused predominantly on pregnant women (Kelpin et al., 2019). While a variety of screening measures for prenatal alcohol use are available (Burns et al., 2010; Svikis & Reid-Quiñones, 2003), the T-ACE is among those most

often used to identify prenatal drinking (Sarkar et al., 2010). Developed by Sokol and colleagues (1989), the T-ACE retains 3 CAGE items (A, C and E), but replaces the G (feeling guilty) item with a question about Tolerance (How many drinks does it take to make you feel high?). In scoring the T-ACE, two points are assigned for the tolerance items when a woman reports needing more than two drinks to feel the intoxicating effects of alcohol or to feel "high." The T-ACE demonstrates good sensitivity and specificity for detecting risk consumption and AUD and has been used to screen for problem drinking in non-pregnant women as well (Kelpin et al., 2019).

Assessing Alcohol Quantity and Frequency

Anther strategy for identifying individuals at risk for alcohol related problems is to examine quantity and frequency of use patterns. Patients with regular alcohol or other substance use are more likely to present for care with chronic and acute medical problems than patients who are not regular drinkers or drug users (Gupman et al., 2002). The past few decades have seen a growing literature describing a plethora of measures focused on quantity and frequency (QF) of alcohol use with psychometric evaluation across diverse study samples (Sobell & Sobell, 2004). QF methods provide reliable and valid information about total consumption (quantity) and number (frequency) of drinking days (McKenna et al., 2018). All QF measures calculate "average" or "typical" consumption patterns (e.g., "How many days *on average*—in a specific time interval (e.g., per week, per month)—did you drink alcohol, and when you drank alcohol, *on average* how many beers did you drink?"). Most QF methods repeat these questions for each major alcoholic beverage type (i.e., beer, wine, distilled spirits) and then sum across beverage types (Sobell & Sobell, 2004). QF methods provide a quick and easy estimate of total amount consumed or total number of drinking days, but often do not describe the variability of a

respondent's drinking (i.e., proportion of drinking occasions in which different numbers of drinks (e.g., 1–2, 5–9, ≥ 10) were consumed or different types of beverages were combined (e.g., beer and whiskey on the same day) (Room, 1990).

The Graduated–Frequency (GF) Measure (Clark & Midanik 1982; Midanik 1994) was developed in response to this criticism. The GF Measure asks respondents to report the frequency of their drinking for different levels of drinking (e.g., 1–2 drinks or 3–4 drinks; highest level is most ever consumed) in a specific time interval for combined beverage types (Sobell & Sobell, 2004). The usual format consists of a question about largest amount followed by a question of frequency of use of applicable quantities falling below the maximum amount consumed (Del Boca & Darkes, 2003).

Because a wide range of drinking patterns (e.g., daily drinking (i.e., alcohol use every day for a specified period) to sporadic heavy/problem drinking) are common in clinical populations (Wu et al., 2017), accurate assessment of such drinking is important. Unfortunately, QF methods to capture such drinking days (e.g., the Volume–Pattern Index and the GF Measure) result is a longer administration time. Recently, with the increased emphasis on brief methods to identify those at risk for alcohol problems, the QF approach has shifted to one or two question instruments (McNeely et al., 2015a; McNeely et al., 2015b; Saitz et al., 2014b; Seale et al., 2006; Smith et al.,2009). Brief QF items, or single-item screening questions (SISQs) typically take the following form, "How many times in the past year have you had X or more drinks in a day?" (where X is 5 for men and 4 for women), with an answer over one designated as positive (Smith et al., 2009; McNeely et al., 2015b), an indication of binge drinking (NIAAA, 2017). However, a positive answer has a sensitivity ranging from 0.34 to 0.89 [95% CI range, 0.25 to 0.92] for male participants and 0.28 to 0.91 [95% CI range, 0.21 to 0.93] for female participants for detecting

unhealthy alcohol use (O'Connor et al., 2018), but an even lower specificity (67-74%) for detecting current alcohol use disorder (McNeely et al., 2015b; Smith et al., 2009). Nonetheless, these one or two item instruments have been shown to offer clinicians something practical and easier to implement (McNeely et al., 2015a; McNeely et al., 2015b; Smith et al., 2009; Strobbe, 2014). The single-question question QF screen is another instrument of choice recommended to quickly identify unhealthy alcohol use in primary care settings (USPSTF, 2018), however, administering an additional instrument (e.g., AUDIT, AUDIT-C) is essential for risk stratification and to guide care (McNeely et al., 2015a).

Timeline-Follow-Back (TLFB)

Originally designed to measure alcohol consumption in problem drinkers, the TLFB has long been the "gold standard." Developed by Sobell and Sobell (1992), the TLFB combines characteristics of a semi-structured interview with that of a brief QF screener to accurately collect information about behaviors. The TLFB is typically administered by trained staff and is facilitated using memory aids, such as calendars and anchor dates. It can be administered by computer, telephone (Pedersen et al., 2012; Sobell et al., 1996), or as a paper-and-pencil survey.

The TLFB retrospectively tracks daily use over a reference period ranging from 30 to 360 days and across many variables (e.g., number of days of low- and high-risk drinking, number of days abstinent, days to relapse, mean drinks per drinking day, mean drinks per week, longest continuous abstinence period). This allows several dimensions of a person's alcohol use to be separately examined: (a) variability (i.e., scatter); (b) pattern (i.e., shape); and (c) extent of use (i.e., elevation; how much), thus providing different and more precise information on individual use levels than indirect estimations (e.g., QF methods) (Sobell & Sobell, 1996; Sobell & Sobell, 2004; Sobell & Sobell, 2000).

The TLFB has been used extensively in the research literature and has been found to have high test-retest reliability, with coefficients ranging from .79 to .96 over 30- to 90-day follow-up periods across a range of drinking populations (Sobell & Sobell, 2004). Many studies have compared QF methods to the TLFB for drinking and have concluded that the TLFB for drinking is superior to other QF methods in its ability to collect useful data (Lemmens et al., 1992; O'Hare et al., 1991; Saunders & Conigrave, 1990) and accurately categorize drinking levels (i.e., heavy drinker, moderate drinker) (Flegal, 1990).

In terms of clinical applicability, the TLFB has been shown to have good psychometric characteristics, but the length of administration (e.g., 15 minutes to complete the TLFB for 90 days and approximately 30 minutes for 12 months) and required training makes it impractical for medical settings such as primary care clinics (Fiellin et al., 2013). Further, some have found that the TLFB strengths may lie more in obtaining an overall summary of drinking patterns, rather than the number of drinks on a specific date (Hoeppner et al., 2010; Searles et al., 2000).

Drug Use and Prescription Drug Misuse Screening and Assessment

DAST

In contrast to alcohol, fewer standardized measures are available (Smith et al., 2010). One measure is the DAST (Drug Abuse Screening Test) (Skinner, 1982), which consists of ten direct obvious questions about drug use and related problems that yields a quantitative score reflecting the severity of drug abuse. Since its inception, the DAST has been used as a clinical and research tool, though it was developed in addiction treatment rather than primary care settings (Smith et al, 2010). The DAST has moderate to high levels of test-retest, interitem, and item-total reliabilities as well as moderate to high levels of validity, sensitivity, and specificity

(McCann et al., 2000), especially for adults in general medical settings at risk for substance use disorders (Yudko et al., 2007). The DAST is available in both 20-item and 10-item formats.

DUDIT

Another measure, the DUDIT (Drug Use Disorders Identification Test) (Berman et al., 2005), an 11-item self-report questionnaire developed to screen individuals for drug problems, was developed in Sweden. Like the DAST and analogous to the AUDIT, the DUDIT assesses an individual's illicit drug use and related consequences over the past year. Specifically, the DUDIT collects data in the following areas: (a) QF of drug use, (b) drug-related problems, and (c) drug dependence symptoms. Validated in both general and clinical populations (e.g., outpatient and residential treatment), the DUDIT was found to be a psychometrically instrument with high convergent validity (r = .85) when compared with the DAST-10, with a Cronbach's alpha of .94. Also, when using the optimal cut-off score of 8, the DUDIT had sensitivity and specificity scores of .90 and .85, respectively (Hildebrand, 2015). A criticism of the DUDIT is that it has not been validated for applicability in general medical settings where the incidence of patients with SUD is typically lower than it is in specialty SUD treatment settings.

SoDU

The Screen of Drug Use (SoDU) is a misnomer, as it was validated in a cross-sectional study of 1283 veterans in primary care to detect drug use disorders (e.g., illicit drug use or prescription drug for non-medical purposes), and not use. The SoDU includes two questions, "How many days in the past 12 months have you used drugs other than alcohol?" (positive if 7 or more and skip the next item) and "How many days in the past 12 months have you used drugs more than you meant to?" (2 or more days meets criteria). The two-question screening instrument was 92.31% sensitive and 92.87% specific (Tiet et al., 2015). Although the SoDU

appears suitable as a screening instrument for drug use disorders, clinical personnel would need to engage patients in secondary-stage screening or other validated, detailed questioners to obtain additional data on drug use.

Similar to single-item alcohol screeners, recommendations for instruments to be administered in primary care settings include the use of a validated SISQs for illicit drug use, "How many times in the past year have you used an illegal drug or used a prescription drug for non-medical reasons?" (McKneely et al., 2015b; Smith et al., 2010), followed by a validated self-report questionnaire (Ghitza et al., 2013).

Combined Alcohol and Drug Use Measures

With polysubstance use as the norm, more screening measures now focus on screening to identify alcohol and/or drug use and quantify risk factors that are associated with actual or potential substance use disorders (Samet et al., 2007). However, assessment measures of this type tend to be quite long, require a significant amount of time to administer (e.g., 5-15 minutes of face-to-face interaction), and involve complex scoring systems.

Several structured interview-based assessment measures collect QF data and focus on both drug and/or alcohol-related problems. One such measure is the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) (WHO, 2002). The ASSIST is a structured interview consisting of 8 questions covering alcohol and use of 10 psychoactive substances. Items cover lifetime use and frequency of use in the past-3 months, as well as various problems associated with the use of these substances. Responses produce a score and subsequent substance-specific risk stratification for each substance (e.g., low-moderate- or high-risk category) (WHO, 2002). Numerous studies have demonstrated good concurrent, construct,

discriminative, and predictive validity (WHO, 2002), as well as test-retest reliability of the ASSIST (Ali et al., 2002; Humeniuk et al., 2008; McNelly et al., 2014; Newcombe et al., 2005).

The length and complexity of the ASSIST have hindered its implementation, especially in primary care settings. To further facilitate adoption, Ali and colleagues (2013) developed a short version of the ASSIST, termed the ASSIST-Lite, based on factor and item-response theory analyses of pooled data from previous ASSIST validation studies. This shortened version shows promising results in identifying problem substance users (sensitivity range: 0.8–1.0; specificity range: 0.7–0.8); however, additional research is needed to establish its clinical utility.

In addition to alcohol, TLFB methods have subsequently been used to collect data on a host of other behaviors (e.g., episodes of violence, gambling behavior, and vocational activity) (Caetano et al., 2012; Svikis et al., 2012), including other drug use. Similar to the Alcohol-TLFB's operating characteristics, the TLFB can be used to obtain information on individual drug use levels more precisely than other indirect or direct estimations (Robinson et al.,2014). The TLFB has been evaluated and used with the following addictive behaviors: nicotine (Shiffman, 2009), cannabis (Norberg et al., 2012), methamphetamine (Halkitis et al., 2009), opiates (Raistrick et al., 1994), prescription drugs (Sellers et al., 1990), cocaine (Ehrman & Robbins, 1994), and polysubstance users (Staines et al., 2001). The TLFB has shown strong test-retest reliability across several substances (Fals-Stewart et al., 2000; Robinson et al., 2014) and diverse populations (Carey et al., 2004; Fals-Stewart et al., 2000; Levy et al., 2004; Sacks et al., 2003). The TLFB has also shown strong concurrent validity using biological measures (Hjorthøj et al., 2012), other interviewer-administered instruments (Dennis et al., 2004; Fals-Stewart et al., 2000) and self-report instruments (Fals-Stewart et al., 2000; Sacks et al., 2003).

As mentioned previously, a major challenge to combined screening for specific substances is provider burden, which refers to the skills and time demanded of the measure. A relatively simple procedure that addresses this problem is the CAGE test adapted to include drugs (CAGE-AID). The CAGE-AID (Brown & Rounds, 1995) was found to be more sensitive but less specific than the CAGE (Brown & Rounds, 1995). This easy-to-use four-item test nevertheless requires further questioning if the patient scores positive. In addition, a brief 4-item patient self-administered Substance Use Brief Screen (SUBS) tool for primary care has been developed and validated. Analyses demonstrated high sensitivity and moderate-to-high specificity for detection of unhealthy use of tobacco, alcohol, and other drugs and indicated good discrimination (AUC 0.74-0.97) for all substance classes (McNeely et al., 2015a).

Computer Surveys/Computer-Directed Substance Use Screening and Assessment

To maximize reliability and validity and facilitate scoring (McCaul & Wand, 2017; Richter & Johnson, 2001), several instruments are available for computer administration. The second edition of *Assessing Alcohol Problems: A Guide for Clinicians and Researchers*, published by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) (Allen & Wilson, 2003) lists 28 computerized alcohol instruments, 21 of which are self-administered. Also, many drug use screening measures, such as the DAST and DUDIT, have been adapted for computer administration.

Furthermore, several instruments can be administered as Audio Computer-Assisted Interviews (ACASI). Computerized-directed measures demonstrate comparable validity to traditional interview formats. For example, an ACASI version of the AUDIT was shown to be feasible, acceptable to patients, and equally good as an interviewer- or pencil-and-paper self-administered version at detecting problem drinking among English speaking patients (Butler,

Chiauzzi et al., 2003). In addition, an ACASI version of the ASSIST was developed and produced scores similar to those obtained by the interview-administered ASSIST, demonstrating good psychometrics to measure substance use in primary care patients (Kumar et al., 2016; McNeely et al., 2014a; McNeely et al., 2016a; Wolff & Shi, 2015).

Recently, McNeely and colleagues (2016b) developed the Tobacco, Alcohol, Prescription medication, and other Substance use (TAPS) tool, a two-stage brief instrument for substance use screening in primary care settings. This ACASI (or interviewer-administered) tool consists of a 4-item screen (TAPS-1) asking about frequency of tobacco, alcohol, illicit drugs, and nonmedical use of prescription drugs, followed by a brief assessment of past 3-month problem use (TAPS-2) to generate a substance-specific risk level for individuals. The TAPS tool was compared to a computerized version of the Composite International Diagnostic Interview (CIDI) Substance Abuse Module (Cottler, 2000; Cottler et al., 1996; Robins et al., 1988).

At a cutoff score of 1+ for problem use the TAPS Tool had sensitivity 0.93 (95% CI 0.90–0.95) and specificity 0.87 (95% CI 0.85–0.89) for tobacco, and sensitivity 0.74 (95% CI 0.70–0.78), specificity 0.79 (95% CI 0.76–0.81) for alcohol. For problem use of illicit and prescription drugs, sensitivity ranged from 0.82 (95% CI 0.76–0.87) for marijuana to 0.63 (95% CI 0.47–0.78) for sedatives, and specificity was 0.93 or higher. For identifying any SUD, sensitivity was lower, but a score of 2+ greatly increased the likelihood of having a SUD. In general, this screening tool demonstrated acceptable sensitivity and specificity for identifying primary care patients with problem substance use, with potential to also detect alcohol, tobacco, and marijuana use disorders (McNeely et al., 2016b).

While the interviewer-administered and self-administered versions of the TAPS Tool performed similarly, when the tool's first-stage screening component (TAPS-1) was validated in

primary care patients, disclosure rates for prescription medication misuse were 50% higher with the self-administered forma (Gryczynski, et al., 2017). A possible advantage of self-administered formats (particularly ACASI formats) is that participants are more reluctant to disclose substance use in a face-to-face interview compared to computerized tools (Butler et al., 2009). Self-administered computer tools also have the potential to facilitate substance use screening in busy medical clinics and potentially streamline screening results into electronic medical records (Gryczynski, et al., 2017; McNeely et al., 2016b).

Methods of Enhancing Validity of Self-Reported Substance Use

There are notable limitations to the validity of self-report data of substance use. One such limitation is the underreporting of use. The under-reporting of substance use can be due to several factors, including poor recall, protection of personal data (confidential or anonymous) in research design and clinical practice, and social desirability bias (responding inaccurately to appear favorably) (McKenna et al., 2018; Rosenbaum et al., 2006).

Social desirability bias may result in underreporting because substance use is often illegal or socially proscribed, stigmatizing, and potentially embarrassing to report (Andreae et al., 2016; Copeland, 1997; Corrigan et al., 2016; Digiusto & Treloar, 2007; Keyes et al., 2010; Kulesza et al.,2013; Mojtabai & Crum, 2013; Schomerus et al., 2010; Tourangeau & Yan, 2007). Research has consistently found higher rates of underreporting of illegal drug use as compared to those for licit substances like alcohol and tobacco use (Gelberg et al., 2015). Further, despite relatively uniform rates of substance abuse among racial and ethnic populations (SAMHSA, 2015), members of racial and ethnic minority groups are perhaps most adversely affected by stigma and are most likely to experience barriers that affect treatment services and outcomes for SUDs (Acevedo et al., 2018; Geurrero et al., 2013). Quantifying differences across race/ethnicity in

drug and alcohol use following screenings requires further investigation (Babor et al., 2007; Sahker et al., 2019).

The TLFB method places emphasis on reducing forgetting and social desirability bias as a source of under-reporting. As mentioned previously, in a TLFB administration, interviewers work actively to address forgetting by facilitating recall events. The face-to-face, semi-structured format establishes trust and rapport and often mitigates social desirability bias. Despite these advantages, however, respondents often minimize their substance use reports to present themselves in a socially desirable manner (Richter & Johnson, 2001). This is consistent with other literature suggesting that respondents report higher rates of drug and alcohol use with a self-administered questionnaire than a face-to-face, provider-administered interview (Rosenbaum et al., 2006).

Regarding protection of personal data – When collected for research purposes, information can be confidential or anonymous. Anonymous data are collected without personal identifiers that can link the information to the participant. When data is collected and held confidentially, the assessors and research staff can identify subjects and may have access to their data. When there are risks associated with disclosure of sensitive information, concerns about the confidentiality of data can impact disclosure of use (Sanker et al., 2003). For example, patients express concerns reporting their substance use in medical records and how that could potentially impact their job, insurance payments for medical care, or the care they receive from their providers (McNeely et al., 2018).

Research suggests that disclosure of stigmatized behaviors, especially on self-administered questionnaires, is enhanced when participants respond anonymously instead of confidentially. While many studies contrast anonymity with confidentiality across a range of

stigmatized health-related behaviors (e.g., HIV, mental health, psychosocial problems) (Bing et al., 2001; Bowling, 2005; Durant et al., 2002; Kessler et al., 2005; Kim et al., 2008; Ong & Weiss, 2000; Richman et al., 1999; Simões & Bastos, 2004; Simões et al., 2006), the number of studies focusing exclusively on alcohol and drug use is limited (Hormes et al., 2012; Richter & Johnson, 2001). Furthermore, methodological concerns arise given a relative lack of agreement over the notion about confidentiality and/or anonymity. For example, some researchers treat these as two exclusive constructs (e.g., Ong & Weiss, 2000; Scott, 2005) while others question whether anonymity and confidentiality are the same (Rogelberg et al., 2006).

Several studies have focused on computerized modes of administration to counteract substance use-related stigma and address confidentiality concerns, demonstrating that computerized self-administered instruments generate higher rates of self-reported substance use (Beck et al.,2014; Delker et al., 2016; Lessler et al., 2000; McNeely et al., 2016a; NIMH, 2008; Perlis et al., 2004; van Griensven et al., 2006). Investigators posit that computerized data collection improves the accuracy of reports due to the provision of anonymity (Joinson, 2001; Newman et al., 2002) and enhanced perceived safety of disclosure (Brandimarte et al., 2012; Spear et al., 2016). Also, computerized-directed tools may yield health benefits for underserved populations by enhancing patient engagement, improving implementation of clinical guidelines, and inform patient care for those most vulnerable who have higher substance abuse and health risks.

The notion that greater anonymity in computerized surveys results in higher self-disclosure of sensitive behaviors, including substance use, is highlighted in a 2015 meta-analysis (Gnambs & Kaspar, 2015). While evaluating disclosure of sensitive behaviors across self-administered paper–and–pencil and computerized survey modes in non-clinical settings,

computerized assessments resulted in significantly ($p < .05$) higher self-disclosure than paper-and-pencil modes. This overall effect was replicated for several subgroups of different types of sensitive behaviors, including various forms of substance use ($\Omega = 1.17$). In particular, the use of heroin or cocaine resulted in larger differences in prevalence rates across survey modes as compared to less sensitive behaviors (predicted $\Omega = 1.43$), such as smoking or the consumption of alcoholic beverages (Gnambs & Kaspar, 2015). While research suggests increased self-disclosure of sensitive behaviors in computerized surveys, future research on privacy perceptions and survey modes on self-disclosure of substance use is highly warranted.

Adjuncts to Self-Report

Since the assessment and evaluation of substance use are largely dependent on self-reports, under-reporting is a significant threat to internal study validity. To test the validity of self-reported substance use, both clinicians and researchers have recommended the use of a self-report measure in combination with biological testing (Center for Substance Abuse Treatment, 1997; Cone, 1997; Donovan et al., 2012), especially in high-risk populations and those with high threats of stigmatization (Clark et al., 2016).

Substance use testing can be completed on several biological matrices including urine, blood, hair, saliva, sweat, nails (toe and finger), and meconium (Moeller et al., 2017). Urine drug testing is the most well-established and supported biological matrix for detection of drug use in clinical settings (Moeller et al., 2008; Stefanidou et al., 2010). As such, it is the most commonly obtained specimen for drug testing, and to improve the accuracy of the assessments of drug use. A significant advantage of urine drug tests over other biological tests is the ability to conduct Point of Care Testing (POCT) in office settings. POCT allows for immediate results on site, allowing providers to review results with the patient in real-time.

Regarding alcohol, alcohol levels can be estimated through collection and analysis of breath, blood, and urine. Alcohol biomarkers include direct and indirect biomarkers (SAMHSA, 2012; Jastrzębska et al., 2016). Indirect biomarkers (CDT, LFT & MCV) identify alcohol's effects on organ systems or body chemistry. In contrast, Phosphatidylethanol (Peth) is a direct biomarker, meaning that it is only formed after someone has consumed alcohol. Phosphatidylethanol testing, known as Peth testing, is a highly reliable blood test allowing the detection of chronic excessive alcohol abuse over the previous 3-4 weeks. With a sensitivity and accuracy rate of over 99%, it is being widely adopted as a replacement to CDT, LFT & MCV testing which offers up to a 77% sensitivity rate (Viel et al., 2012). Although not recommended for routine screening, these methods have proven useful to increase the validity of self-report information in general (Babor et al., 2007).

Research comparing biochemical verification results with self-reported data have primarily been conducted in substance use treatment settings and have demonstrated high rates of concordance (Basurto et al., 2009; Jain et al., 2006; Jain et al., 2013; Melnikov et al., 2009; Schuler et al., 2009) while other studies have yielded mixed results (Clark, et al., 2016; Cone, 2012; Neale and Robertson, 2003). In general, a variety of self-report measures were used across these studies.

The TLFB has gained popularity in cross sectional and prospective studies of drug use, and a 2012 systematic review and meta-analysis concluded that the TLFB is a valid measure in detecting illicit substance use in populations with substance use disorders (Hjorthøj et al., 2012). However, Nordeck et al. (2020) found mixed results in a study comparing the TLFB and oral fluid testing to detect substance use in adult primary care patients. Specifically, they found that marijuana use had higher detection using TLFB self-report, while cocaine, prescription opioids,

and heroin had higher detection using oral fluid testing. Additional research is needed in order to contribute to the literature on identifying substance use in primary care settings and to evaluate the concordance of TLFB compared to biological testing.

Statement of Problem

While much attention has been focused on brief interventions (BIs) and their effects on substance use, screening to identify persons at risk for heavy/problems alcohol or drug use has received much less attention. This is unfortunate as screening is an integral component of SBIRT, as it identifies the individuals likely to benefit from education or BI.

The ability of screening tools to identify persons at risk for heavy/problem alcohol or other drug use varies greatly and can be impacted by types of questions asked and mode of administration. While differences between screening tools and modes of administration have been found, with self-administration and provisions of anonymity resulting in greater disclosure, the assumption in substance use research is that structured interview-administered substance use screening and assessment approaches are most valid. Research exploring the role of consent condition on disclosure in the context of computer-mediated and self-administered data collection in clinical settings is limited. Also, the accuracy and efficacy of screening, especially for illicit drug use and prescription drug misuse, necessitates more research given the insufficient evidence of the clinical utility of these instruments when applied in primary care practice settings.

Moreover, previous studies suggest that in general, demographic characteristics such as age and gender (Beck et al., 2014; Bjarnason & Adalbjarnardottir, 2000; Dolezal et al., 2012; Ledgerwood et al., 2008; Vigil-Colet et al., 2015; Welte & Russell, 1993), race (Johnson & Fendrich, 2005; Ledgerwood et al., 2008; White et al., 2014), mental health-related conditions

(Baggio et al., 2015), and socioeconomic status (Welte & Russell, 1993) may be related to differential reporting of sensitive health behaviors. Whether these characteristics differentially affect reporting of substance use frequency in interviewer-administered vs. computer-mediated screening and assessments among primary care patients is largely unknown. Further, additional research is needed among those with comorbid SUD and medical disorders to be able to inform integrated screening and care-coordinating efforts (Freeman et al., 2014; Walter et al., 2017).

The present study compares self-reports of alcohol, illicit drug use, and prescription drug misuse across an anonymous, computer-administered health screen (CAHS), confidential, computer-administered research assessment (CARA), and confidential, interviewer-administered research assessment (IARA) in males and females with heavy/problem substance use identified through an urban primary care clinic. This study also compares self-reports of recent substance use to urine drug assay findings. Data was part of a randomized clinical trial (RCT) evaluating a brief computerized Screening, Brief Intervention, and Referral to Treatment. This novel approach of comparing computerized anonymous and confidential surveys to a confidential, interviewer-administered assessment across alcohol and illicit drug use as well as prescription drug misuse will offer insight on the most parsimonious methods to collect the most sensitive and accurate data possible about substance use and related behaviors.

Aims and Hypotheses

The present study had four primary aims:

I. Aim 1: To compare rates of participant self-disclosure of alcohol and illicit drug use and prescription drug misuse obtained by an anonymous, computer-administered health screen to those obtained by a confidential, computerized assessment and interview-administered research assessment. Four hypotheses were tested:

a. Hypothesis 1: Participants will report more recent binge drinking days (past 30) on the anonymous, computer-administered screener (CAHS) compared to the confidential, interviewer-administered assessment (IARA).

b. Hypothesis 2: Participants will report consuming, on average, more drinks per week over the past 30 days by anonymous, computer-administered screener (CAHS) compared to the confidential, computerized research assessment (CARA). The study also hypothesized that participants will report consuming, on average, more drinks per week by CAHS compared to the interview-administered assessment (IARA).

c. Hypothesis 3: Participants will report, on average, more days of drug use per week over the past 30 days by anonymous, computer-administered screener (CAHS) compared to confidential, computerized assessment (CARA). The study also hypothesized that participants will report, on average, more days of drug use per week by CARA compared to the interview-administered assessment (IARA).

d. Hypothesis 4: Participants will report, on average, more days of prescription drug misuse per week by anonymous, computer-administered screener (CAHS) compared to confidential, computerized assessment (CARA). The study also hypothesized that participants will report, on average, more days of drug use per week by CAHS compared to the interview-administered assessment (IARA).

II. Aim 2: Compare endorsement of alcohol and/or drug use related problems between the anonymous, computer-administered screener and confidential, computerized assessment.

a. Hypothesis 1: Participants will be more likely to score positive for problematic substance use on the anonymous, computer-administered screener compared to the confidential, computerized assessment.

III. Aim 3: To compare agreement between self-report of recent drug use and biological measures (i.e., urinalysis).

a. Hypothesis 1: Proportion of illicit drug use and non-medical prescription drug use will be higher by urine assay compared to self-report measure.

IV. Aim 4: To identify correlates (e.g., age, gender, race, marital status, years of education, ethnicity, number of medical conditions, mental health-related conditions, etc.) of self-disclosure of any alcohol and other drug use as well as substance use per week in the past 30-days.

Methods

Participants

Present student participants were drawn from a parent randomized, controlled clinical trial (RCT) of SBIRT for heavy/problem alcohol or other drug use in a primary care patient sample. The study was approved by Virginia Commonwealth University's Institutional Review Board under "Project COMP: A Randomized Clinical Trial," protocol number HM13196 and all participants provided informed consent. Study recruitment occurred from November 2010 to December 2013, until the total sample was achieved.

Parent Study RCT

Phase 1

In the parent study, Phase 1 participants were N=4,552 patients who completed an anonymous computer-administered health survey (CAHS) that focused on demographics and

general health and health risk behaviors including sleep, mood, diet and exercise, smoking and alcohol and other drug use.

Phase 2

Based on CAHS data who screened at risk for heavy problem alcohol or drug use, patients were invited to participate in the 4-arm RCT. The alcohol and drug use items to determine RCT eligibility were embedded within the larger health survey to minimize any stigma associated with completing a screener about substance use.

RCT Inclusion Criteria

For the parent RCT was as follows: between 18 – 70 years of age; not pregnant (by self-report); residing in the clinic catchment area; able to speak and understand English; not enrolled in substance abuse treatment (inpatient, residential, outpatient, methadone maintenance, therapeutic community), and screened positive for heavy/problem alcohol and/or drug use. Further details about heavy/problem substance use are described below in the methods for the current study.

Current Study

Design and Procedures

Patients were recruited from Virginia Commonwealth University Health System (VCUHS) (City of Richmond, Virginia and 33 surrounding counties) primary care and gynecology clinic waiting areas. They were approached by research assistances (Ras) who invited them to participate in CAHS. Those who expressed interest were escorted to a private area adjacent to the clinic waiting room. After obtaining informed verbal consent participants completed the 10-minute CAHS (Health Cheq) via Table computer.

Random Assignment

Eligible patients who consented to the RCT (N=713) were randomized to one of four study arms, with stratification. Randomization to one of the four study conditions was determined by the computer, using participant survey data and internal algorithms, with stratification by gender; race (Caucasian or minority), and primary substance type (drug or alcohol). Current study participants were *N* = 540 individuals who completed the CAHS, met criteria for the RCT, and were randomized to either the CA, CACI, CATI arms of the study. Those in the TAU group were excluded as they did not participate in further baseline assessment means central to the current study hypotheses.

Group 1. Treatment as usual (TAU) (standard care true control): Participants randomized to this arm completed no further research assessments until 3 & 6-months follow-ups.

Group 2. Confidential, computerized assessment (CA) intervention only: Participants randomized to this arm were asked more detailed questions about their alcohol and drug use. Assessment items included additional standardized assessment measures (e.g., AUDIT, Babor et al., 2001), more detailed information about quantity and frequency of use with detailed characterization (using timeline-follow-back (TLFB) methods (Sobell & Sobell, 1992) in the past 30 days. Drug items looked both generally (across all classes of drugs combined) and more specifically at drug classes prevalent in the City of Richmond (i.e., marijuana, cocaine, heroin/other opiates, sedative-hypnotics). Additional alcohol and other drug use questions focused on episodes of heavy use, problems associated with use (e.g., craving, preoccupation, loss of control, arguments with spouse, legal problems), efforts to quit or reduce use, family history of problems and personal history of substance use treatment (e.g., AA/NA, outpatient, residential, detoxification).

Group 3. Confidential, computerized assessment plus computer-delivered BI (CACI): Participants randomized to this arm completed the CA plus a computer-delivered BI that utilized a Motivation Enhancement System (MES). The MES included traditional motivationally-oriented intervention components: feedback, pros and cons, willingness to change, and optional goal-setting.

Group 4. Confidential, computerized assessment plus therapist-delivered BI intervention (CATI): Participants randomized to this arm completed the CA plus a single-session MI counseling session that focuses specifically on the same elements described above.

After the computer and TLFB assessments, CA, CACI, and CATI participants were asked to provide a urine sample assayed for 6 drug classes (cocaine, amphetamines, marijuana (THC), opiates, methadone, and sedative/hypnotics). Baseline visit participation took between 5-40 additional minutes (depending on group assignment). Patients were compensated for completing the assessments, with gift cards, and had the potential to earn $150 for completing all scheduled visits.

Computerized Software Platform

The computerized screening and assessment for the proposed study were developed using the Computerized Intervention Authoring System (CIAS), an authoring tool that allows creation or editing of internet-delivered interventions without the need of a programmer. The software features a high-quality synthetic text to speech engine that reads all questions and speaks aloud to the participant (using headphones); synchronous interactivity, natural language reflections, branching logic, a clean user interface, tailored SMS, and the ability to easily incorporate specific images, graphs, figures, text, or videos. CIAS was designed to be consistent with research from the Human-Computer Interaction literature suggesting that ethopoeia (the extent to

which software embodies lifelike attributes such as a voice, image, or personality) is related to greater engagement and better outcomes (e.g., Appel et al., 2012; Brave et al., 2005; Mumm & Mutlu, 2011); CIAS, therefore, uses an interactive and emotive three-dimensional narrator with multiple engaging animations. This narrator reads and speaks aloud and functions as an engaging guide throughout the intervention.

For the parent and current study, the CAHS, computer-directed assessment, and computer-directed BI utilized CIAS. A mobile three-dimensional cartoon character, Peedy the Parrot, capable of over 50 specific animated actions (e.g., smile, wave, read a message, express concern, etc.) does the "talking" for the entire program. Peedy the Parrot acted as a narrator, reading each item aloud for the participant, guiding them throughout the survey, and providing occasional comic relief. Pleasing and relevant graphics changed with each screen to maintain interest. Participants could use either the touchscreen or keyboard/mouse to proceed through the survey, and participants listened to Peedy via headphones to ensure privacy. All answers were provided by choosing responses from a list or by touching a visual analog scale. The program did not require reading literacy.

Alcohol and Other Drug Use Screening and Assessment Measures

All screening and assessment measures are summarized in Table 2. Please see Appendix A for the TLFB used in the present study. Below is a detailed description of the alcohol, illicit drug, and prescription drug misuse screening and assessment measures.

Table 2. *Baseline Screening and Assessment Measures.*

Domain	Measure	Anonymous, Computer-Administered Health Screen	Confidential, Computer-Administered Research Assessment	Confidential, interviewer-Administered Research Assessment
Demographic Characteristics	Age, race, ethnicity, gender, marital status, education, employment, income, insurance coverage	**F**		
	Q/F of typical recent (past 30 days) use	**F**		
	Episodes of binge drinking	**F**		
Alcohol Use	AUDIT		**P**	
	Q/F of lifetime (heaviest) alcohol use		**F**	
	Q/F of past 3 months use		**F**	
	Timeline-Follow-Back (Q/F)			**F**
	CAGE (men)	**F**		
	T-ACE (Women)	**F**		
Alcohol-related Problems	ASSIST (loss of control, problems, cravings)		**P**	
	Efforts to quit or reduce drinking, family history of alcohol problems and personal history of alcohol treatment		**P**	
	Brief Mast		**P**	
Illicit Drug Use	Frequency of recent (past 30 days) use (not drug-specific)	**F**		
	Frequency of recent (past 30 days) regular (3+x's/week) use (drug specific)	**F**	**F**	
	Frequency of lifetime use (drug specific)		**P**	
	Frequency of use (drug specific) – past 3 months		**F**	
	Peer influence on use		**P**	

	Timeline-Follow-Back (Q/F)			**F**
Illicit Drug Use-related Problems	CAGE-DRUG (men and women)	**F**		
	DAST		**P**	
	Family history of drug problems	**P**		
	ASSIST (loss of control, problems, cravings)		**P**	
Prescription Drug Misuse	Frequency of recent (past 30 days) misuse	**F**		
	Lifetime use (drug specific)		**P**	
	Episodes of heavy use (drug specific)		**P**	
	Frequency of past 3-month use (drug specific)		**P**	
	Reasons for misuse and source of drug		**F**	
	Peer influence on drug use		**P**	
	Timeline-Follow-Back (Q/F)			**F**
Prescription Drug Misuse-related Problems	ASSIST (loss of control, problems, cravings)		**P**	
	DAST		**P**	
Tobacco Use	Quantity and frequency of recent and lifetime (heaviest ever) use	**F**		
	Fagerstrom Tobacco Dependence Questionnaire (FTD)	P		
Physical and Emotional Health	General health questions	P		
Depression	K10	F		
Anxiety	K10	F		
Partner Violence	PVS	F		

F=Full Assessment
P=Partial Assessment

CAHS Tools for Alcohol Use

Screening measures included the 4-item CAGE (Ewing, 1998) for men and the 4-item T-ACE (Sokol et al., 1989) and items from the 10-item Brief MAST (Pokorny et al., 1972). Problem alcohol use represents >2 score on the CAGE (for men) and T-ACE (for women). Participants were asked about quantity and frequency of typical recent (past-30 days) and binge drinking. This period of assessment is congruent with the TLFB assessment period (past 30 days) and thus enabled a comparable evaluation.

CAHS Tools for Illicit Drug Use & Prescription Drug Misuse

Screening included the standardized screening tool, the 4-item CAGE-AID, adapted for use of illicit drugs (CAGE-DRUG), with a cut-off of ≥ 1 (for both men and women) indicating problem substance use (Brown & Rounds, 1995), and subitems from the DAST (Skinner, 1982). Also, participants were shown a computer screen listing various classes of drugs (modified from the CIDI drug module) (Uston et al., 1997) and were asked to check all those they have used regularly (3+ days/week) in the past 30 days. The list included marijuana, cocaine, stimulants, inhalants, heroin, and hallucinogens. For those drugs endorsed, they were asked about their frequency of use in the past 30 days, including how many days they used as well as their average use per week. Various classes of prescription drugs (narcotics/analgesics, sedative-hypnotics) were included along with a list of common medications or drugs found within each class. Subsequent items asked about any category checked yes and focused on the frequency of recent use (past 30 days), use by prescription, and variations from such use (e.g., using greater amounts or using for longer periods than prescribed). Problems associated with prescription drug misuse was not assessed by the CAHS. Similar to CAHS tools for alcohol use, the past 30-day illicit drug use and/or prescription drug misuse period of assessment is congruent with the TLFB

assessment period (past 30 days) and thus enabled a comparable evaluation. Family history of drug use was also assessed.

CARA Tools for Alcohol Use

Assessment included subitems from additional standardized measures (e.g., AUDIT) (Babor et al., 2001), information about quantity and frequency of use, including lifetime (heaviest) alcohol and their average use per week in the past 3-months. Further assessment included subitems from the ASSIST (WHO, 2002) that referred to the past 3-month period and asked three questions focused on problems associated with alcohol use (e.g., loss of control, arguments with partner/family members related to drinking, craving or needing a drink). Response categories for these items were: almost all the time, sometimes, rarely, and not at all. Additional questions focused on efforts to quit or reduce drinking, family history of alcohol problems, and personal history of alcohol treatment (AA, outpatient, residential, detoxification). Participant perception of treatment was also surveyed.

CARA Tools Illicit Drug Use & Prescription Drug Misuse

For illicit drug use and prescription drug misuse, additional assessment focused on past 3-month use, including how many days they used as well as their average use per week. Questions also focused on regular use, lifetime use, and episodes of heavy use/misuse. Items looked both generally (across all classes of drugs combined) and more specifically at drug classes prevalent in the City of Richmond (i.e., marijuana, cocaine, heroin/other opiates, sedative-hypnotics). Past 3-month problematic use (e.g., loss of control, fights with partner/family members related to use, and cravings) was assessed using three subitems from the ASSIST (WHO, 2002) and focused separately for each illicit drug (i.e., marijuana, cocaine, amphetamines, inhalants, heroin, and hallucinogens) and/or prescription drug (i.e., sleep medicine, sedative and anxiety medicine,

stimulant medicine, and pain medicine) the participant self-reported using or misusing. Response categories for these items were: almost all the time, sometimes, rarely, and not at all. Additional questions focused on efforts to quit or reduce use/misuse, family history of drug problems, and personal history of drug treatment (NA, outpatient, residential, methadone maintenance, detoxification). Participant perceptions of such treatment experiences were assessed. Reasons for using drugs and peer influence on drug use were also assessed.

IARA for Alcohol Use and Other Drug Use.

Detailed information about quantity and frequency of alcohol use in the past 30 days was obtained using Timeline-Follow-Back (TLFB) methods (Sobell & Sobell, 1992), and detailed characterization of both past 30-day use of illicit drugs as well as prescription drug misuse was assessed using TLFB procedures.

Urine Drug Assay.

Participants were asked to provide a urine sample that was assayed for 6 drug classes (marijuana (THC), cocaine, amphetamines, opiates, oxycontin, and benzodiazepines). Information was collected about prescription medications that participants reported taking medically as prescribed and urine drug screens (UDS) for those medications were *not* counted as positive.

In regard to the classification of UDS samples, positive opiate results could encompass both illicit substances (e.g., heroin) and non-medical prescription use (e.g., opiates, oxycontin) categories due to similar metabolic processes. For the present study, positive tests for opiates with self-reported heroin use (i.e., the heroin metabolite, 6-Monoacetylmorphine) absent of self-reported non-medical prescription opiate use were classified as heroin-positive; tests positive for other opioids (codeine, oxycontin, oxycodone, hydrocodone) absent of self-reported heroin use

were classified as positive for non-medical prescription opioid use. Also, positive tests for non-medical amphetamine and methamphetamine use were classified as amphetamine positive.

Additional Baseline Assessment Measures

Tobacco. Items from the NSDUH (SAMHSA, 2006) were used to briefly summarize quantity and frequency of recent and lifetime (heaviest ever) use. The items quantified lifetime cigarette smoking with questions such as: "Which statement best describes your smoking behavior over your lifetime?" with 3 response options (1 = 100 or more cigarettes and 3 = I have never smoked cigarettes). Also included were questions about lifetime and recent use (age of onset of use and daily smoking, years of daily smoking, efforts to quit, current interest in quitting or reducing tobacco use) (SAMHSA, 2006). Fagerstrom Test for Nicotine Dependence (Heatherton et al., 1991) was also used. These items assessed a range of current smoking behaviors including the number of cigarettes smoked per day, the amount of time after waking in the morning to first cigarette, and which cigarette would be the most difficult to give up. The Fagerstrom Test for Nicotine Dependence has demonstrated high reliability, as well as validity when using cotinine as a criterion variable (Pomerleau et al., 1994).

Physical and Emotional Health. Participants were presented with a range of medical conditions related to addiction (e.g., chronic obstructive pulmonary disease (COPD), arthritis, hepatitis, liver disease, pancreatitis) and asked whether they had ever received a diagnosis for any of the listed conditions. They were also asked about the reason for their current medical visit (e.g., yearly check-up, new health problems, ongoing health problems) and, in general, how would they rate their overall health on a 5-point scale (1 = excellent and 5 = poor).

Kessler Psychological Distress Scale (K10). Level of psychological distress was assessed using the Kessler Psychological Distress Scale (Kessler & Mroczek, 1992; Kessler et al., 2002).

This short dimensional measure includes 10 items to measure participants' emotional state in the past 30 days, which are scored on a 5-point scale (1 = none of the time and 5 = all of the time) and identifies levels of distress. Scores of the 10 items are then summed, with higher scores indicating higher levels of psychological distress.

Sleep Behavior. Sleep behavior was assessed using items from the Insomnia Severity Index (Bastian et al., 2001). Participants were asked to rate if they had difficulty falling asleep, staying asleep, or problems waking up too early on a 5-point scale (1 = none and 5 = very severe). The Insomnia Severity Index corresponds with the DSM-IV criteria for insomnia and measures perceptions of symptom severity, distress, and daytime impairment. The diagnostic validity of the measure has been well established in distinguishing individuals diagnosed with primary insomnia from good sleeper controls (sensitivity, 94%; specificity 94%) (Smith & Wegener, 2003). The use of prescription drugs for sleep was also assessed with an item from the Pittsburgh Sleep Quality Index (PSQI) asking participants if they took any medication to help them sleep in the past 30 days, and if so, how often did they take this medication as a sleep aid on a 4 point scale (1 = daily and 4 = less than twice per week) (Buysse et al., 1989). This measure has also demonstrated diagnostic validity in distinguishing good and poor sleepers (sensitivity, 89.6%; specificity, 86.5%).

Partner Violence Screen (PVS). The 3-item Partner Violence Screen was included to assess two dimensions of partner violence: physical violence and perceived safety (Feldhaus et al., 1997). Physical violence was assessed by asking, "Have you been hit, kicked, punched, or otherwise hurt by someone within the past year?" with yes/no option. Two questions assessed perceived safety by asking, "Do you feel safe in your current relationship?" and "Is there a partner from a previous relationship who is making you feel unsafe now?" with yes/no options.

Three studies have assessed the sensitivity and specificity of the PVS (sensitivity, 35%-71%; specificity, 80%-94%) (Feldhaus et al., 1997; Mills et al., 2006; MacMillan et al., 2009), and although it has demonstrated a wide range of sensitivity, it serves as a brief screen for identifying partner violence in the primary care setting.

Defining Heavy/Problem Substance Use

To qualify for the RCT and the current study, participants had to meet criteria for heavy/problem alcohol and/or drug use. These were defined as follows:

Heavy/problem alcohol use was defined as either: 1) 1) CAGE alcohol score ≥ 2 (men) or T-ACE score ≥ 2 (women) and self-reported consumption of > 14 drinks/week (men) or > 7 drinks/week (women) in the past 30 days; or 2) Self-reported consumption of 5 or more drinks (men) or 4 or more drinks (women) on at least two occasions in the past 30 days.

Heavy/problem drug use was defined as 1) CAGE drug score ≥ 1 and recent drug use (past 30 days); or 2) illicit drug use 2 or more days/week (past 30 days); or 3) misuse of prescribed medications (e.g., taking more than prescribed, using someone else's prescription, getting medications from more than one health provider) on at least 2 occasions in the past 30 days.

Current Study Variables

Variables for the current study included demographic variables (e.g., age, gender, race, ethnicity, employment, and marital status), alcohol and other drug use variables, and UDS results. The alcohol and other drug use variables were carefully selected and/or created out of existing alcohol and other drug use screening and assessment measures based on domains to be studied as well as psychometric properties. Table 3 lists the current study variables and the crosswalk between variables in the CAHS, CARA, and IARA. Past 30-day alcohol and binge

drinking as well as other drug use were also included to provide direct comparisons between the anonymous, computer-administered health screen and confidential, interview-administered research assessment, as past 30-day use was directly assessed by both methods. However, for past 30-day binge drinking and prescription drug misuse the anonymous, computer-administered health screen offered broader response options (e.g., 11-20 days, 21-29 days) and were reclassified as 15 and 25 days, respectively. Past 30-day alcohol and other drug use on the confidential, computer-administered research assessment was calculated by taking the mean of the self-reported average drinking days and/or other drug use (each drug was asked separately and included all those the participant reported using) per week in the past three months and multiplying by 4.29.

Table 3. *Crosswalk Between Screening and Assessment Study Variables.*

	Measure/Question				
Domain	**Anonymous, Computer-Administered Health Screen**		**Confidential, Computer-Administered Research Assessment**		**Confidential, interviewer-Administered Research Assessment**
Alcohol Use – Weekly (7/days)	During the past 30 days, on average, how many drinks did you have each week? (asked separately for both females and males)	**X**	In an average week in the past 3 months, how many days per week did you typically drink? + In the past 3 months, on the days when you did drink, how many drinks did you have?	**X**	Quantity of weekly use in past 30 days
Binge Use	During the past 30 days, how many times have you had 4/5 or more drinks per occasion? (asked		**N/A**		Number of binge drinking days in the past 30 days

	separately for both females and males)				
Illicit Drug Use – Weekly (7/days)	During the past 30 days, on how many days each week did you use recreational drugs? (not drug-specific)	**X**	In an average week in the past 3 months, how many days per week did you typically use xxx? (drug specific)	**X**	Frequency of weekly drug use (drug specific)
Prescription Drug Misuse – Weekly (7/days)	In the past 30 days, on how many days have you: Taken more pills, more often, etc. (not drug-specific)	**X**	In an average week in the past 3 months, ("thinking about misuse") how many days per week did you typically use xxx medicine? (drug specific)	**X**	Frequency of weekly misuse (drug specific)
Alcohol-related Problems (males)	CAGE	**X**	ASSIST-3 items		N/A
Alcohol-related Problems (females)	T-ACE	**X**	ASSIST-3 items		N/A
Drug-related problems (males and females)	CAGE-DRUG	**X**	ASSIST-3 items		N/A

The anonymous, computer-administered health survey directly asked for average weekly use for alcohol (i.e., quantity and frequency) and illicit drug use in the past 30 days. For prescription drug misuse, mean days of misuse per week was calculated by dividing past 30-day

use by 4.29. Days of alcohol and other drug use in the past 30 days were divided by 4.29 to calculate the average weekly use on the interviewer-administered research assessment. Mean days per week in the past three months was used to calculate average weekly use for alcohol and other drug use on the confidential, computer-administered research assessment. Drinks per week in the past 30-days were grouped the same across each measure (0 drinks, 1-2 drinks, 3-5 drinks, 6-9 drinks, 10-15 drinks, 16-20 drinks, and 21 or more drinks).

The T-ACE was modified by assigning only 1 point to the tolerance question while keeping positivity set at 2 or more of the possible 4 questions. The T-ACE was originally developed for a prenatal population to detect risk of fetal alcohol syndrome where a positive response to the tolerance question was sufficient for further inquiry. However, in a primary care clinic population, the modest increase in sensitivity resulting from lowering the threshold of positivity is not offset by the considerable drop in specificity (i.e., false positives) (McQuade, Levy, Yanek, Davis, & Liepman, 2000).

The 3-Question ASSIST (ASSIST-3) questions for alcohol read, “In the past 3 months, how often did you experience a loss of control while you were drinking?” (ASSIST Q1); “In the past 3 months, how often did your drinking lead to problems like fights with family and friends?” (ASSIST Q2); and “In the past 3 months, how often did you have strong cravings for alcohol?” (ASSIST Q3). For illicit drug use, ASSIST questions read, “In the past 3 months, how often did you experience a loss of control while you were using (_____)?” (ASSIST Q1); “In the past 3 months, how often did your (_____) use lead to problems like fights with family and friends?” (ASSIST Q2); and “In the past 3 months, how often did you have strong cravings for (_____)?” (ASSIST Q3).

Similar to Tiet and colleagues (2015), the present study consolidated the ASSIST-3 questions that assessed each category of illicit drug separately (i.e., marijuana, cocaine, heroin, amphetamines, inhalants, and hallucinogens) to assess all illicit drugs as a combined question. For every illicit drug the participant self-reported using regularly, ASSIST-3 responses were calculated to create a total ASSIST-3 score. The highest score (i.e., most severe value) for each question was summed. Therefore, if a participant reported high-risk for marijuana on one question but only moderate risk for cocaine, only the high-risk score was summed.

Data Analysis

Data Preparation and Missing Data

Analyses for all study aims were computed with SPSS software, version 27.0 (SPSS, Armonk, NY). Prior to analysis, means, standard deviations and 95% confidence intervals (or medians and inter-quartile ranges) were estimated for each continuous variable, while frequencies, proportions, and 95% confidence intervals were computed for each categorical variable. Data were also examined for normality of distribution and the presence of outliers.

Missing data was limited within the CAHS and CARA since the screening and assessment questions were delivered via CIAS with complex skip patterns based on participant responses. Missing data were also limited within the IARA given the face-to-face and semi-structured format. Therefore, analyses were conducted to include all participants who were randomized to CA, CACI, CATI groups. The data was treated as is, with any case with missing values excluded from the analyses.

Primary Analyses

Aim 1: To address the first study aim comparing rates of recent (past 30-day) alcohol use, including binge drinking, illicit drug use (including marijuana, cocaine, methamphetamine,

heroin), and non-medical use of prescription drugs (including opioids, stimulants, and sedative-hypnotics) across groups (e.g., CAHS by CARA, CARA by IARA, CAHS by IARA), sensitivity and specificity, positive predictive values (PPV), negative predictive values (NPV), and chance-corrected Cohen's kappa coefficient (**κ**) statistics were calculated. Kappa values ≤ 0 indicate no agreement, 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement (Cohen, 1960). It has been suggested that k values of 0.50 or 0.60 are indicative of "acceptable" concordance (Grove et al., 1981). The interviewer-administered TLFB is the accepted "gold standard" against which the anonymous screener and computerized assessment (when applicable) are compared in order to reflect testing new modalities against the status quo.

Interrater reliability across modes of administration for frequency of use (days during the past 30) as well as self-reported drinks per week (7/days) and frequency of weekly (7/days) illicit drug use and prescription drug misuse during the past 30 days was calculated using intraclass correlation coefficient (ICC) (Shrout & Fleiss, 1979) and their 95% confident intervals. The ICC (2,1) was based on a single-rater/measurement, consistency, 2-way random-effects model. Because we have randomly selected our raters (i.e., anonymous computer-administered screener, computerized assessment, interview-administered TLFB) from a larger population of raters with similar characteristics, a 2-way random-effects model was the model of choice. This 2-way random-effects model allowed us to generalize our reliability results to any raters who possess the same characteristics as the selected raters in this study. The selection of single-rater/measurement was used because we plan to use the measurement from a single rater (interviewer-administered TLFB) as the basis of the actual measurement; therefore, "single rater" type was selected even though the reliability experiment involves 2 or more raters. Because the

CARA does offer a direct comparison of weekly alcohol and other drug use against the CAHS and IARA, the definition of relationship considered is "consistency" versus "absolute agreement." Based on the 95% confident interval of the ICC estimate, values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability (Koo & Li, 2016; Shrout & Fleiss, 1979).

Hypothesis 1. The first hypothesis predicted that participants will report more binge drinking days in the past month (30 days) on the anonymous, computer-administered screener compared to the confidential, interviewer-administered assessment. To test this hypothesis, binge drinking days were compared for the two groups using paired sample t-tests. The effect size for the paired-samples t-test was calculated using Cohen's d (Cohen. 1998).

Hypothesis 2. The second hypothesis predicted that participants will report consuming, on average, more drinks per week over the past 30 days by anonymous, computer-administered screener (CAHS) compared to the confidential, computerized assessment (CARA). The study also hypothesized that participants will report consuming, on average, more drinks per week by CAHS compared to the interview-administered assessment (IARA). To test this hypothesis, drinks per week were compared for the three groups using paired sample t-tests. The effect size for the paired-samples t-test was calculated using Cohen's d.

Hypothesis 3. The third hypothesis predicted that participants will report, on average, more days of drug use per week over the past 30 days by anonymous, computer-administered screener (CAHS) compared to confidential, computerized assessment (CARA). The study also hypothesized that participants will report, on average, more days of drug use per week by CAHS compared to the interview-administered assessment (IARA). To test this hypothesis, days of drug

use per week were compared for the three groups using paired sample t-tests. The effect size for the paired-samples t-test was calculated using Cohen's d.

Hypothesis 4. The fourth hypothesis predicted that participants will report, on average, more days of prescription drug misuse per week by anonymous, computer-administered screener (CAHS) compared to confidential, computerized assessment (CARA). The study also hypothesized that participants will report, on average, more days of drug use per week by CAHS compared to the interview-administered assessment (IARA). To test this hypothesis, rates of prescription drunk misuse per week were compared for the three groups using paired sample t-tests. The effect size for the paired-samples t-test was calculated using Cohen's d.

Aim 2: Aim 2 compared endorsement of alcohol and/or drug use related problems between the CAHS (CAGE, T-ACE, CAGE-DRUG) and CARA (ASSIST-3).

First, descriptive statistics were calculated for problems related to alcohol use identified by the CAGE, T-ACE, and ASSIST-3 items. Descriptive statistics were also calculated for problems related to illicit drug use identified by the CAGE-DRUG as well as lifetime history of regular (3+ days/week) illicit drug use and subsequent ASSIST-3 scores. The CAGE-DRUG and ASSIST-3 items for illicit drug use were analyzed separately for males and females.

Two-tailed independent t-tests were used to compare ASSIST-3 scores for participants endorsing each of the items (yes/no) of the CAGE, T-ACE, and CAGE-DRUG questions. The CAHS (i.e., CAGE-DRUG) did not measure problems associated with prescription drug misuse.

For alcohol, the sample size was N = 159 for males. Excluded were N = 25 with missing data due to a programming error. N = 26 males who reported no alcohol use on both the CAHS and CARA and therefore were not asked both CAGE and ASSIST-3 items. For T-ACE, the sample size was N = 285 for this analysis. Excluded were women (N = 45) who denied alcohol

use on both the CAHS and CARA and therefore were not asked both CAGE and ASSIST-3 items.

For CAGE-DRUG (males), the sample size was N = 144. Excluded were N = 66 males who denied drug use on both surveys and therefore were not asked both CAGE and ASSIST-3 items. For CAGE-DRUG (females), the sample size was N = 220, and excluded were N = 110 women who reported no drug use on both the CAHS and CARA and therefore were not asked both CAGE and ASSIST-3 items.

The ASSIST-3 total scores for alcohol and ASSIST-3 scores for illicit drug use were investigated for its ability to identify problematic alcohol and illicit drug use on the CAGE, T-ACE, and CAGE-DRUG. The CAGE, T-ACE, and CAGE-DRUG was used as the reference measure. These tools have been widely used in primary care settings to assess substance use-related problems (Brown & Rounds, 1995; Lanier & Ko, 2008; Sarkar et al., 2010; Tan et al., 2018; USPSTF, 2018). Substance use-related problems were defined using the standard threshold of scoring positive on two or more CAGE/T-ACE items and scoring positive on one more CAGE-DRUG items (Steinbauer et al., 1998).

Sensitivity, specificity, and the positive and negative predictive values (PPV and NPV) were calculated. The sensitivity and the specificity represent the proportion of participants whose CAGE, T-ACE, CAGE-DRUG score were correctly identified by the ASSIST-3 cut-points and the proportion of patients who do not have a problematic CAGE, T-ACE, CAGE-DRUG score and who have a negative ASSIST-3 total score (i.e., below the cut-points) respectively. The PPV represents the proportion of patients above the cut-points on the ASSIST-3 who had a problematic score on the CAGE, T-ACE, CAGE-DRUG, while the NPV shows the proportion of

patients who test negative on the ASSIST-3 who do not have a problematic score on the CAGE, T-ACE, CAGE-DRUG.

To generate a cut-point on the ASSIST-3 total score, binary logistic regression analyses and receiver operating characteristic (ROC) analysis were performed. The area under each curve (AUC) was examined (Hanley & McNeil, 1982) and the cut-point for the ASSIST-3 total score was based on maximizing the AUC. Excellent discrimination was considered an AUC of >.90, very good discrimination was considered an AUC of 0.8-0.9, acceptable discrimination was considered 0.7-0.8, poor discrimination was considered 0.6-0.7, while an AUC of 0.5-0.6 was considered poor discrimination (Meyers et al., 2017).

Hypothesis 1. Participants will be more likely to score positive for problematic substance use on the anonymous, computer-administered screener (i.e., CAGE, T-ACE, CAGE-DRUG) compared to the confidential, computerized assessment (i.e., ASSIST-3). This hypothesis was tested using chi-square analyses.

Aim 3: Aim 3 compared agreement between self-reports of recent drug use (marijuana, cocaine, heroin) or non-medical prescription drug use (opiates, benzodiazepines, amphetamines) on the TLFB and biological measures (i.e., urinalysis). The TLFB recorded only non-medical use of prescription medications. Those who did not provide a urine drug screen (UDS) (N = 26) or who's illicit drug misuse data were missing (N =49) were excluded from the analysis. Also, type of prescription drug used was not initially assessed on the IARA due to parent study design. As such, these participants (N = 109) were also excluded from the analysis.

Substances reported on the TLFB were matched with those detected on the urinalysis, and a binary variable was created for each substance, separately for both the TLFB and UDS.

Comparisons were made separately for self-reported non-medical prescription opiate use and heroin use as well as those with both prescription opiate and heroin use.

Self-reported use on the TLFB and UDS results were compared for overall agreement. Participants were classified as concordant (positive) for each respective substance if the individual endorsed use for the identified substance on the IARA (i.e., the TLFB) and had a positive UDS result. Likewise, participants were classified as concordant (negative) if they reported no recent substance use and had a negative result on the UDS for each respective substance. Participants were classified discordant if the individual was TLFB positive only (i.e., self-reported use in combination with a negative UDS test) or UDS positive only (i.e., no self-reported use in combination with a positive UDS test). Sensitivity, specificity, positive predictive values (PPV), and negative predictive values (NPV) were also calculated.

The kappa (**k**) statistic (Cohen (1960)) was used to examine the concordance of self-report and urinalysis results. While it has been suggested that k values of 0.50 or 0.60 are indicative of "acceptable" concordance (Grove et al., 1981), a combination of factors could potentially inflate the value of k. For example, underestimation/under-reporting of substance use by self-report, administering only one urinalysis tests during the 30-day interval covered by the TLFB, and potential for false-negatives with the UDS testing technique could potentially increase the apparent agreement between the two measures. Thus, in order to adjust for these factors, the conditional k statistic (cond. k) was also computed (Bishop et al., 1975). The cond. k indicates the degree to which self-reports agree with a positive urinalysis result and has been used to examine validity of self-reported drug use in several studies (Chermack et al., 2000; Sherman and Bigelow, 1992; Zanis et al., 1994).

Hypothesis 1: Proportion of illicit drug use and non-medical prescription drug use will be higher by urine assay compared to self-report measure. This hypothesis was tested using chi-square analyses.

Aim 4: Lastly, univariate logistic regression was used to identify correlates of self-disclosure of any alcohol and other drug use as well as substance use per week in the past 30-days, sampling from psychosocial and health disparity domains. Variables included: age; gender; race (e.g., White/Minority); ethnicity; marital status (married/unmarried); years of education (≤ 12 yrs./< 12 yrs.); insurance coverage (private/government or state); employment status; anxiety; depression; psychological distress (moderate/severe) as measured by K10 total score (Andrews & Slade, 2001); sleep behavior (quality of sleep; taking sleep aids); and number of medical conditions (0/≥ 2). Significance was set at 0.05 for all univariate analyses; however, any variables reaching a significance level of <0.20, along with basic demographic variables, were subsequently included in a multivariate logistic regression model to examine their effect in combination. When multiple items from the same domain reached significance, only one predictor was selected for inclusion in the multivariate analysis to avoid multicollinearity.

A multivariable logistic regression model, with backward elimination, was used to identify the most parsimonious model of correlates of self-reported substance use. The final model was achieved by eliminating covariates, one by one, that were not significant at the 0.05 level. The Hossmer Lemeshow test was used to check goodness-of-fit of the logistic regression, as well as the R-squared value to determine how much of the variance was explained by the model.

Results

Participant Flow through Study

The consort diagram (see Figure 1) summarizes participant flow through the parent RCT and current study, as per the Standards of Reporting Trials guideline (Altman et al., 2001). A

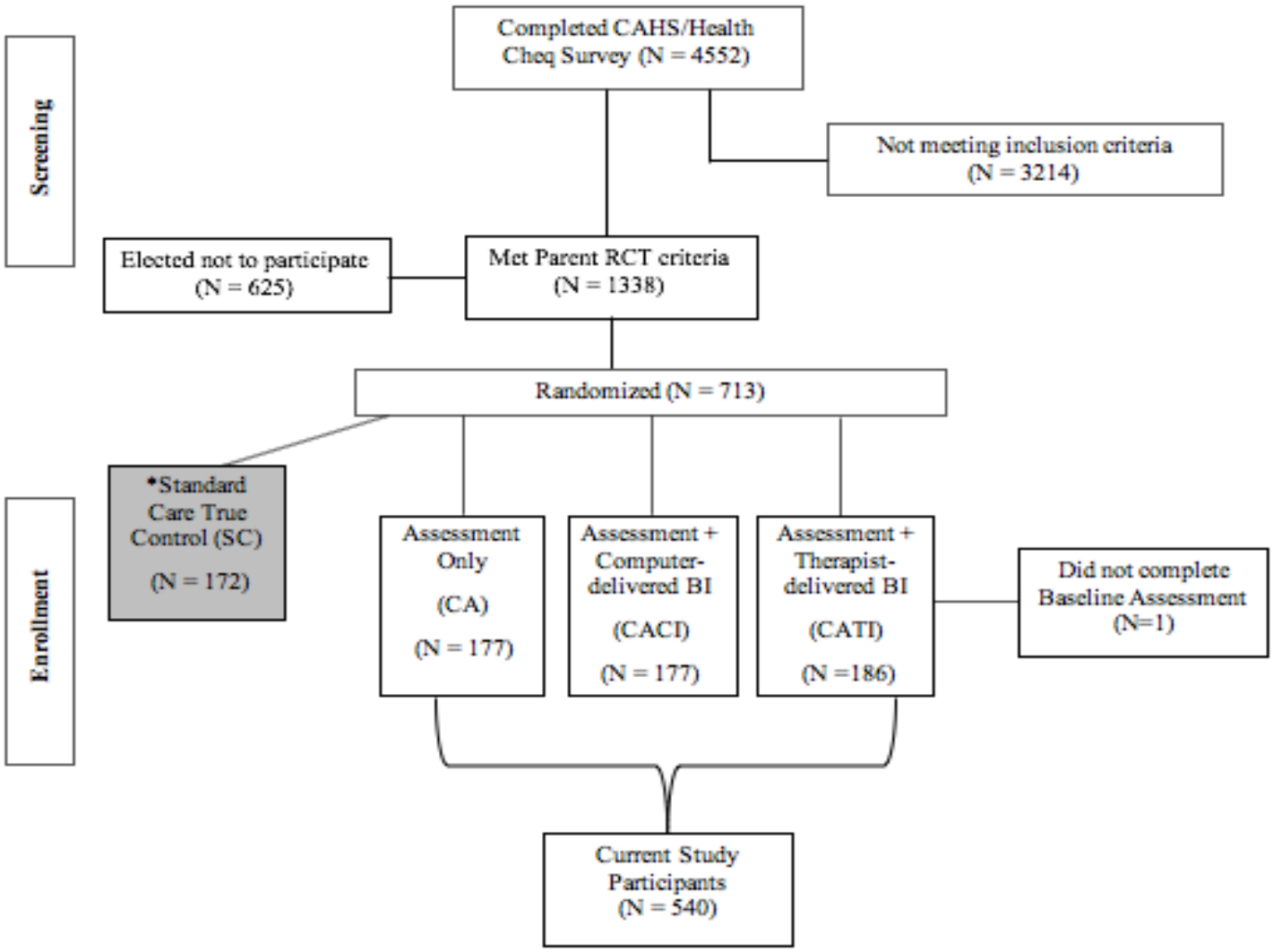

Figure 1. CONSORT Diagram of Study Recruitment and Enrollment.
*Did not complete all baseline measures.

total of N = 4552 primary care patients completed the CAHS/Health Cheq Survey. Of those, N = 3214 did not meet inclusion criteria; of those who met parent RCT criteria, N = 625 elected not

to participate. A total of N = 713 provided informed consent and were randomized. A total of N = 172 were randomized to the Standard Care True Control group and did not complete all baseline measures. One participant (N=1) did not complete the three baseline assessments for the present study and was removed from analyses, totaling a sample size of N = 540 for the current study.

Demographics

Table 4 summarizes demographic characteristics for those who completed the anonymous, computer-administered health screen (CAHS) and went on to complete the

Table 4. *Participant Demographic Characteristics* (*N=540).*

Variable	Percent (N)	Mean ± SD
Age (yrs.)		45.14 (11.01)
Gender – Male	39% (210)	
Race		
Black/African American	78% (421)	
White	18% (98)	
Other	4% (21)	
Ethnicity		
Hispanic	3% (18)	
Employment*		
Full Time	12% (49)	
Part Time	13% (56)	
On Disability	20% (83)	
Retired	3% (13)	
Unemployed	49% (206)	
Student	3% (13)	
Marital Status*		
Single/Never Married	45% (187)	
In a Relationship	21% (86)	
Married	13% (53)	
Divorced/Separated	20% (82)	
Widowed	3% (12)	

*N=421 sub-sample for these variables.
SD =standard deviation

confidential, computer-administered research assessment (CARA), and confidential, interviewer-administered research assessment (IARA) (*N*=540). Almost two-thirds of the participants were female (61%). Mean age was 45.14 years (*SD* = 11.01) and the majority identified their race as Black/African American (78%). Nearly half reported being unemployed (49%) and a similar percentage were single/never married (45%).

Overall Substance Use

The alcohol and other drug use characteristics for the overall sample are summarized in Table 5. As required for study eligibility, all participants met criteria for heavy/problem alcohol and/or drug use.

Table 5. *Baseline Substance Use for the Overall Sample* (*N=540).*

Substance Use and Problems Variable	CAHS % (N)	CARA % (N)	IARA % (N)
Alcohol Use			
Any Use (Past 30-Day)	86% (465)	59% (320)	79% (424)
Any Binge Use (Past 30-Day)	71% (382)	--	43% (232)
Alcohol Problems			
T-ACE + (females N=330)	39% (130)	--	--
CAGE + (males N=185)*	49% (90)	--	--
Drug Use			
Any Illicit Drug Use (Past 30 Days)	37% (198)	53% (284)	47% (253)
Any Prescription Drug Misuse (Past 30 Days)	27% (144)	27% (145)	27% (145)
Type of Prescription Drug Misuse			
Higher Dosage than Prescribed	15% (83)	--	--
More Often than Prescribed	12% (62)	--	--
Using Someone Else's Prescription	13% (69)	--	--
Obtaining Same Prescription from Multiple Physicians	2% (10)	--	--
Drug Problems			
CAGE-DRUG +	39% (208)	--	--

*N=25 males excluded due to computer branching error.

AIM 1: Rates of Any Recent Alcohol and Other Drug Use by Assessment Method (CAHS, CARA, or IARA)

Any Recent (Past 30-Day) Alcohol Use. As shown in Table 6, rates of recent (past 30-day) alcohol use by self-report ranged from 59% by confidential, computerized research assessment (CARA), to 79% by confidential, interviewer-administered research assessment (IARA), and 86% by anonymous, computer-administered health screen (CAHS). When the IARA (i.e., TLFB) was considered the gold standard, sensitivity for detecting alcohol use ranged from 0.70 for the CARA to 0.97 for the CAHS. The specificity of CAHS was moderate (0.54), with lower specificity for the CARA (0.21). Positive Predictive Values (PPV) were high for both the CAHS (0.89) and CARA (0.93) while Negative Predictive Values (NPV) were mixed, ranging between 0.42 and 0.84 for the CARA and CAHS, respectively. There was moderate agreement between both the CAHS and IARA ($\kappa = 0.59$) and fair agreement between the CARA and IARA ($\kappa = 0.37$). When the CARA was considered the gold standard, CAHS sensitivity was excellent (0.99) but specificity was fair (0.32). PPV was moderate at 0.68 and NPV was high at 0.95, with fair agreement between the two assessments ($\kappa = 0.34$).

Table 6. *Any Recent (Past 30-Day) Alcohol Use by Assessment Measure (N=540).*

	Any Recent (Past 30-Day) Alcohol Use					
	CAHS	**IARA**	CAHS	**CARA**	CARA	**IARA**
Any Recent Use % (N)	86% (465)	79% (424)	86% (465)	59% (320)	59% (320)	79% (424)
+ on Both Assessments % (N)	76% (411)		59% (315)		47% (255)	
Sensitivity	0.97		0.99		0.70	
Specificity	0.54		0.32		0.21	
Positive Predictive Value (PPV)	0.89		0.68		0.93	
Negative Predictive Value (NPV)	0.84		0.95		0.42	
Kappa	0.59		0.34		0.37	

Any Recent (Past 30-Day) Binge Drinking. Rates of self-reported binge drinking (past 30-days) for the CAHS and IARA are shown in Table 7. The CARA did not measure this variable. Over two thirds of participants reported recent binge drinking on the CAHS compared to only 43% on the IARA. With the IARA (i.e., TLFB) as the gold standard, sensitivity and specificity for the CAHS was 0.91 and 0.45, respectively. The PPV (0.55) was moderate while the NPV (0.87) was high, and there was overall fair agreement between the CAHS and IARA ($\kappa = 0.33$).

Table 7. *Any Recent (Past 30-Day) Binge Drinking by Assessment Measure (N=540).*

	Any Recent (Past 30-Days) Binge Drinking	
	CAHS	**IARA**
Any Recent Use % (N)	71% (382)	43% (232)
+ on Both Assessments % (N)	39% (211)	
Sensitivity	0.91	
Specificity	0.45	
Positive Predictive Value (PPV)	0.55	
Negative Predictive Value (NPV)	0.87	
Kappa	0.33	

Any Recent (Past 30-Day) Illicit Drug Use. Rates of any recent (past 30-day) illicit drug use ranged from 37% by CAHS to 47% by IARA to 53% by CARA (see Table 8). With the IARA (i.e., TLFB) considered the gold standard, sensitivity for CAHS was 0.71 with higher

Table 8. *Any Recent (Past 30-Day) Illicit Drug Use by Assessment Measure (N=540).*

	Any Recent (Past 30-Day) Illicit Drug Use					
	CAHS	**IARA**	CAHS	**CARA**	CARA	**IARA**
Any Recent Use % (N)	37% (199)	47% (253)	37% (199)	53% (285)	53% (285)	47% (253)
+ on Both Assessments % (N)	34% (180)		34% (184)		42% (224)	
Sensitivity	0.71		0.65		0.89	
Specificity	0.94		0.95		0.79	
Positive Predictive	0.91		0.93		0.79	

Value (PPV)			
Negative Predictive Value (NPV)	0.79	0.71	0.89
Kappa	0.66	0.58	0.67

sensitivity for the CARA at 0.89. Conversely, specificity ranged from 0.79 to 0.94 for the CARA and CAHS, respectively. PPV was high at 0.79 for the CARA and even higher for CAHS (0.91), while NPV for the CAHS was 0.79 and 0.89 for the CARA. Kappa statistic for rater agreement was substantial between both the CAHS and IARA ($\kappa = 0.66$) and CARA and IARA ($\kappa = 0.67$). With CARA as the gold standard, sensitivity for the CAHS was at 0.65 with excellent specificity (0.95), and PPV was high at 0.93 with lower NPV at 0.71. There was moderate agreement ($\kappa = 0.58$) between the two instruments.

Any Recent (Past 30-Day) Prescription Drug Misuse. Table 9 shows rates of any recent (past 30-day) prescription drug misuse across the three assessments. Over one-fourth (27%) of participants reported any recent misuse on each measure. With the IARA (i.e., TLFB)

Table 9. *Any Recent (Past 30-Days) Prescription Drug Misuse by Assessment Measure (N=540).*

	Any Recent (Past 30-Day) Prescription Drug Misuse					
	CAHS	**IARA**	CAHS	**CARA**	CARA	**IARA**
Any Recent Use % (N)	27% (145)	27% (145)	27% (145)	27% (145)	27% (145)	27% (145)
+ on Both Assessments % (N)	18% (98)		16% (83)		17% (93)	
Sensitivity	0.68		0.57		0.64	
Specificity	0.88		0.84		0.87	
Positive Predictive Value (PPV)	0.68		0.58		0.64	
Negative Predictive Value (NPV)	0.88		0.84		0.87	
Kappa	0.56		0.42		0.51	

as the gold-standard, sensitivity ranged from 0.64 for the CARA to 0.68 for the CAHS. Specificity was 0.87 for the CARA and 0.88 for the CAHS. PPV ranged from 0.64 for the CARA

to 0.68 for the CAHS while NPV was high at 0.87 for CARA and 0.88 for CAHS. CAHS sensitivity was slightly lower (0.57) when the CARA was considered the gold standard, with specificity at 0.84. PPV remained moderate (0.58) while NPV remained high (0.84). Kappa statistic demonstrated moderate agreement between the CAHS and IARA ($\kappa = 0.56$), CAHS and CARA ($\kappa = 0.42$), and CARA and IARA ($\kappa = 0.51$).

Frequency of Recent (Past 30-day) Alcohol and Other Drug Use

Frequency of Past 30-Day Binge Drinking. As shown in Table 10, among those reporting any binge use, number of binge drinking days in the past 30 averaged 5.37 (SD = 7.77, median = 2.00, range = 30) on the CAHS and 3.52 on the IARA (SD = 7.10, median = 0.00, range = 30). Participants (N = 171) reported one or more binge drinking days only on the CAHS, while N = 21 reported one-plus days of binge drinking only on the IARA, and N = 211 reported 1+ more days with binge use on both. The remainder (not in this analysis, N = 137) reported no alcohol use on either measure.

Hypothesis 1. Participants will report more recent binge drinking days (past 30) on the anonymous, computer-administered screener (CAHS) compared to the confidential, interviewer-administered assessment (IARA). As shown in Table 10, there was a significant difference in

Table 10. *Days of Recent (Past 30-Day) Binge Drinking by Assessment Measure (N=540).*

Survey Methods	Mean ± SD	Mean Difference	95% CI of the Difference Lower	Upper	*t* (*p*-value)	Cohen's d
CAHS	5.37 ± 7.77					
x		1.85	1.26	2.44	6.15 (< 0.001)*	0.26
IARA	3.52 ± 7.10					

*Denotes a statistically significant t-value ($p < .05$).
SD = standard deviation.

mean binge drinking days for the CAHS compared to the IARA (t(539) = 6.15, $p < 0.001$), 95% CI [1.26, 2.44], Cohen's d = 0.26, with more binge drinking days on the CAHS as compared to the IARA. These results support the hypothesized associations.

Frequency of Past 30-Day Alcohol Use. Table 11 summarizes days of recent alcohol use across the three assessment measures. Among those reporting any alcohol use, mean number of drinking days in the past 30 ranged from 7.97 on the IARA (SD = 9.47, median = 4.00, range = 30) to 8.48 (SD = 9.57, median = 4.29, range = 40) on the CARA, and 10.10 on the CAHS (SD = 10.17, median = 6.00, range = 30). In the CAHS – IARA comparison, N = 53 participants reported number of days of use only on the CAHS, N = 12 reported one-plus days of use on the IARA, and N = 411 reported 1+ more days of use on both. The remainder (not in this analysis, N = 64) reported no alcohol use on either measure. In the CAHS – CARA comparison, N = 149 participants reported number of days of use only on the CAHS, N = 4 reported one-plus days of use only on the CARA, and N = 315 reported 1+ more days of use on both. The remainder (not in this analysis, N = 72) reported no alcohol use on either measure. In the CARA – IARA comparison, N = 24 participants reported number of days of use only on the CARA, N = 128 reported one-plus days of use on the IARA, and N = 296 reported 1+ more days of use on both. The remainder (not in this analysis, N = 92) reported no alcohol use on either measure.

Table 11. *Days of Recent (Past 30-Day) Alcohol Use by Assessment Measure.*

Days (Past 30) of Alcohol Use					
CAHS	**IARA**	CAHS	**CARA**	CARA	**IARA**
Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD
10.10 ± 10.17	7.97 ± 9.47	10.13 ± 10.17	8.48 ± 9.57	8.48 ± 9.57	7.97 ± 9.47

SD = standard deviation.

Frequency of Past 30-Day Illicit Drug Use. Table 12 reflects the number of days participants reported using illicit drugs in the past 30. Average days of drug use ranged from 3.56

(SD = 5.04, median = 0.72, range = 30) on the CARA, to 5.02 (SD = 9.85, median = 0.00, range = 30) on the CAHS, and 6.19 (SD = 10.26, median = 0.00, range = 30) on IARA. In the CAHS – IARA comparison, N = 18 participants reported number of days of use only on the CAHS, N = 73 reported one-plus days of use on the IARA, and N = 180 reported 1+ more days of use on both. The remainder (not in this analysis, N = 269) reported no illicit drug use on either measure. In the CAHS – CARA comparison, N = 14 participants reported number of days of use only on the CAHS, N = 100 reported one-plus days of use on the CARA, and N = 184 reported 1+ more days of use on both. The remainder (not in this analysis, N = 242) reported no illicit drug use on either measure. In the CARA – IARA comparison, N = 60 participants reported number of days of use only on the CARA, N = 29 reported one-plus days of use on the IARA, and N = 224 reported 1+ more days of use on both. The remainder (not in this analysis, N = 227) reported no illicit drug use on either measure.

Table 12. *Recent (Past 30-Day) Illicit Drug Use by Assessment Measure (N=540).*

Days (Past 30) of Illicit Drug Use					
CAHS	**IARA**	CAHS	**CARA**	CARA	**IARA**
Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD
5.02 ± 9.87	6.19 ± 10.26	5.02 ± 9.87	3.56 ± 5.04	3.56 ± 5.04	6.19 ± 10.26

SD = standard deviation.

Frequency of Past 30-Day Prescription Drug Misuse. Table 13 presents the number of days participants reported misusing prescription drugs in the past 30. Average days of drug use ranged from 1.71 on the CAHS (SD = 4.99, median = 0.00, range = 30) to 2.35 on the CARA (SD = 5.23, median = 0.00, range = 30), and 2.95 on the IARA (SD = 7.31, median = 0.00, range = 30). In the CAHS – IARA comparison, N = 46 participants reported number of days of use only on the CAHS, N = 47 reported one-plus days of use on the IARA, and N = 98 reported 1+ more days of use on both. The remainder (not in this analysis, N = 349) reported no prescription

drug misuse on either measure. In the CAHS – CARA comparison, N = 61 participants reported number of days of use only on the CAHS, N = 62 reported one-plus days of

Table 13. *Recent (Past 30-Day) Prescription Drug Misuse by Assessment Measure (N=540).*

Days (Past 30) of Prescription Drug Misuse					
CAHS	**IARA**	CAHS	**CARA**	CARA	**IARA**
Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD
1.71 ± 4.99	2.95 ± 7.31	1.71 ± 4.99	2.35 ± 5.23	2.35 ± 5.23	2.95 ± 7.31

SD = standard deviation.

use on the CARA, and N = 83 reported 1+ more days of use on both. The remainder (not in this analysis, N = 334) reported no prescription drug misuse on either measure. In the CARA – IARA comparison, N = 52 participants reported number of days of use only on the CARA, N = 52 reported one-plus days of use on the IARA, and N = 93 reported 1+ more days of use on both. The remainder (not in this analysis, N = 343) reported no prescription drug misuse on either measure.

Past 30-Day Alcohol and Other Drug Use Reliability Comparisons. ICC estimates of interrater reliability agreement for frequencies of past 30-day alcohol and other drug use across the survey methods are summarized in Table 14. ICC for any recent (past 30-day) alcohol use

Table 14. *Intraclass Correlation Coefficients (ICC) for Past 30-Day Alcohol and Other Drug Use by Assessment Measure (N=540).*

Frequency Variable	Survey Methods	ICC
Days of Recent Alcohol Use	CAHS x IARA	0.75 [0.70, 0.79]
	CAHS x CARA	0.74 [0.69, 0.78]
	CARA x IARA	0.80 [0.76, 0.83]
Days of Recent Binge Drinking	CAHS x IARA	0.72 [0.67, 0.76]
	CAHS x IARA	0.85 [0.83, 0.88)

Days of Recent Illicit Drug Use	CAHS x CARA	0.62 [0.54, 0.66)
	CARA x IARA	0.67 [0.61, 0.72]
Days of Recent Prescription Drug Misuse	CAHS x IARA	0.52 [0.44, 0.60]
	CAHS x CARA	0.43 [0.33, 0.52]
	CARA x IARA	0.53 [0.44, 0.60]

*Values in square brackets indicate the 95% confidence interval for each ICC.

demonstrated moderate interrater reliability between the CAHS and CARA (ICC (2,1) = 0.74) and good reliability for CAHS and IARA (ICC (2,1) = 0.75) and for CARA and IARA (ICC (2,1) = 0.80). ICC also demonstrated moderate reliability for recent (past 30-day binge drinking between the CAHS and IARA (ICC (2,1) = 0.72). ICC for past 30-day illicit drug use indicated moderate reliability between the CAHS and CARA (ICC (2,1) = 0.62 and between the CARA and IARA (ICC (2,1) = 0.67), while ICC demonstrated good reliability between the CAHS and IARA (ICC (2,1) = 0.85). ICC calculations indicated poor agreement between the CAHS and CARA (ICC (2,1) = 0.53), with reliability slightly rising between the CAHS and IARA (ICC (2,1) = 0.52) as well as between the CARA and IARA (ICC (2,1) = 0.53) for past 30-day prescription drug misuse.

Mean Number of Drinks per Week During the Past 30 Days. The previous tables presented days of alcohol and other drug use (past 30 days). The following tables summarize average number of drinks per week of alcohol during the past 30 days as well as days of other drug use per week during the past 30 days. Means, standard deviations, and estimates of interrater reliability (ICC) for the number of self-reported drinks per week (7 days) across the three survey methods are summarized in Table 15. Number of drinks per week averaged from 1.44 on the CARA (SD = 1.90), 2.21 on the CAHS (SD = 1.70), and 2.63 on the IARA (SD = 3.16). ICC demonstrated moderate reliability between the CAHS and CARA (ICC (2,1) = 0.68),

while there was poor reliability between the CAHS and IARA (ICC (2,1) = 0.56) as well as between the CARA and IARA (ICC (2,1) = 0.55). Effect sizes between the CAHS and IARA were small (0.17), with larger effect sizes between the CAHS and CARA (0.42) and between the CARA and IARA (0.46).

Hypothesis 2. Participants will report consuming, on average, more drinks per week over the past 30 days by anonymous, computer-administered screener (CAHS) compared to the confidential, computerized assessment (CARA). The study also hypothesized that participants will report consuming, on average, more drinks per week by CAHS compared to the interview-administered assessment (IARA). As shown in Table 15, there was a significant average difference between the CAHS and IARA (t(539) = -3.50, p = 0.001, 95% CI [-0.66, -0.19], Cohen's d = 0.15), between the CAHS and CARA (t(539) = 9.99, p < 0.001, 95% CI [0.61, 0.91], Cohen's d = 0.43), and between the CARA and IARA (t(539) = -9.50, p < 0.001, 95% CI [-1.43, -0.94], Cohen's d = 0.41). Specifically, participants reported a greater number of drinks

Table 15. *Mean Drinks per Week During the Past 30 Days by Assessment Measure (N=540).*

	Survey Methods	Mean ± SD	Mean Difference	95% CI of the Difference Lower	Upper	*t* (*p*-value)	Cohen's d	ICC
Pair 1	CAHS x IARA	2.21 ± 1.70 2.63 ± 3.16	-0.42	-0.66	-0.19	-3.50 (0.001)*	0.15	0.56 [0.48, 0.63]
Pair 2	CAHS x CARA	2.21 ± 1.70 1.44 ± 1.90	0.76	0.61	0.91	9.99 (<0.001)*	0.43	0.68 [0.62, 0.73]
Pair 3	CARA x IARA	1.44 ± 1.90 2.63 ± 3.16	-1.19	-1.43	-0.94	-9.50 (<0.001)*	0.41	0.55 [0.47, 0.62]

*Denotes a statistically significant t-value (p < .05).
*Values in square brackets indicate the 95% confidence interval for each ICC.

per week on the CAHS compared to the CARA, but greater number of drinks per week were reported on the IARA when compared to both the CAHS and CARA. These results support the hypothesized association between the CAHS and CARA but do not support the hypothesized association between CAHS and IARA.

Illicit Drug Use (Days per Week) During the Past 30 days. Table 16 reflects the average days of drug use per week. Days of drug use per week ranged from 0.83 on the CARA (SD = 1.18) to 1.42 on the CAHS (SD = 2.38), and to 1.44 on the IARA (SD = 2.39). ICC demonstrated good reliability between the CAHS and IARA (ICC (2,1) = 0.86), while there was moderate reliability between the CARA and IARA (ICC (2,1) = 0.61) as well as between the CAHS and CARA (ICC (2,1) = 0.61). Effect sizes across each pair were small.

Hypothesis 3. Participants will report, on average, more days of drug use per week over the past 30 days by anonymous, computer-administered screener (CAHS) compared to confidential, computerized assessment (CARA). The study also hypothesized that participants will report, on average, more days of drug use per week by CARA compared to the

Table 16. *Days of Illicit Drug Use per Week During the Past 30 Days by Assessment Measure (N=540).*

	Survey Methods	Mean ± SD	Mean Difference	95% CI of the Difference Lower	Upper	*t* (*p*-value)	Cohen's d	ICC
Pair 1	CAHS x IARA	1.42 ± 2.38 1.44 ± 2.39	-0.02	-0.16	0.12	-0.32 (0.749)	0.01	0.86 [0.83, 0.88]
Pair 2	CAHS x CARA	1.42 ± 2.38 0.83 ± 1.18	0.59	0.43	0.76	6.93 (< 0.001)*	0.30	0.61 [0.54, 0.67]
Pair 3	CARA x IARA	0.83 ± 1.18 1.44 ± 2.39	-0.61	-0.77	-0.45	-7.60 (< 0.001)*	0.33	0.67 [0.61, 0.72]

*Denotes a statistically significant t-value ($p < .05$).
Values in square brackets indicate the 95% confidence interval for each ICC.

interview-administered assessment (IARA). As shown in Table 16, paired sample t-tests found significant differences between the CAHS and CARA (t(539) = 6.93 $p < 0.001$, CI [0.43, 0.76], Cohen's d = 0.30) and between the CARA and IARA (t(539) = -7.60 $p < 0.001$, CI [-0.77, -0.45], Cohen's d = 0.33). No significant difference was found between the CAHS and IARA comparison (t(539) = -0.32, $p = 0.749$, CI [-0.16, 0.12], Cohen's d = 0.01). Participants reported significantly more days of illicit drug use per week on the IARA and CAHS compared to the CARA. Results do not support the hypothesized association between the CAHS and IARA. As hypothesized, the CAHS found more days of illicit drug use per week than the CARA (t(539) = 6.93 $p < 0.001$, CI [0.43, 0.76], Cohen's d = 0.30).

Prescription Drug Misuse (Days per Week) During the Past 30 Days. Table 17 presents mean days of prescription drug misuse per week for the three administration methods with paired t-tests. Days of drug use per week ranged from less than half a day (Mean = 0.39) on

Table 17. *Days of Prescription Drug Misuse per Week During the Past 30 Days by Assessment Measure (N=540).*

	Survey Methods	Mean ± SD	Mean Difference	95% CI of the Difference Lower	Upper	*t* (*p*-value)	Cohen's d	ICC
Pair 1	CAHS x IARA	0.39 ± 1.16 0.68 ± 1.70	-0.29	0.07	-0.43	-4.08 (< 0.001)*	0.18	0.52 [0.44, 0.60]
Pair 2	CAHS x CARA	0.39 ± 1.16 0.54 ± 1.21	-0.15	-0.27	-0.03	-2.38 (0.018)*	0.11	0.43 [0.33, 0.52]
Pair 3	CARA x IARA	0.54 ± 1.21 0.68 ± 1.70	-0.14	-0.29	-0.00	-2.01 (0.045)*	0.08	0.53 [0.44, 0.60]

*Denotes a statistically significant t-value (p < .05).
*Values in square brackets indicate the 95% confidence interval for each ICC.

the CAHS (SD = 1.16) to 0.54 on the CARA (SD = 1.21), and to 0.68 on the IARA (SD = 1.70). ICC demonstrated poor reliability between the CAHS and CARA (ICC (2,1) = 0.43), while reliability slightly increased between the CAHS and IARA (ICC (2,1) = 0.52) and between the CARA and IARA (ICC (2,1) = 0.53). Effect sizes were small for CARA versus IARA (d = 0.08), for CAHS versus CARA (d = 0.11), and for CAHS versus IARA (d = 0.18).

Hypothesis 4. Participants will report, on average, more days of prescription drug misuse per week by anonymous, computer-administered screener (CAHS) compared to confidential, computerized assessment (CARA). The study also hypothesized that participants will report, on average, more days of drug use per week by CAHS compared to the interview-administered assessment (IARA). As shown in Table 17, there was a significant average difference between the CAHS and IARA (t(539) = -4.08, $p < 0.001$, CI [0.07, -0.43], Cohen's d = 0.18), between the CAHS and CARA (t(539) = -2.38, $p = 0.018$, CI [-0.27, -0.03], Cohen's d = 0.11), and between the CARA and IARA (t(539) = -2.10, $p = 0.045$, CI [-0.29, -0.00], Cohen's d = 0.08). Participants reported, on average, less days of prescription drug misuse per week on the CAHS when compared to both the CARA and IARA, with participants reporting most days of prescription drug misuse on the IARA. Results do not support the hypothesized associations.

Aim 2: Endorsement of Alcohol and/or Other Drug Use Related Problems by CAGE, T-ACE, CAGE-DRUG, and ASSIST-3.

Rates of alcohol-related problems as measured by the CAGE, T-ACE, and ASSIST-3 (e.g., loss of control, fights with partner/family members related to use, and craving) are summarized below (Table 18 and 19). Both males and females were more likely to report problems related to feeling the need to cut down on their drinking on the CAGE/T-ACE and experiencing cravings on the ASSIST-3.

Table 18. *Endorsement of Alcohol-Related Problems by CAGE and ASSIST-3 Items in Males (N = 159).*

CAGE	Yes N (%)	**ASSIST**	Not at all N (%)	Rarely N (%)	Sometimes N (%)	Almost all the time N (%)
Cut down?	110 (69%)	Loss of control?	99 (62%)	24 (15%)	29 (18%)	7 (4%)
Annoyed?	68 (43%)	Problems?	121 (76%)	17 (11%)	16 (10%)	5 (3%)
Guilty?	70 (44%)	Cravings?	85 (54%)	27 (17%)	32 (20%)	15 (9%)
Eye-opener?	69 (43%)					
Summary Score N (%)		Summary Score N (%)				
0. 40 (25%)		0. 72 (45%)	5. 10 (6%)			
1. 29 (18%)		1. 22 (14%)	6. 8 (5%)			
2. 21 (13%)		2. 13 (8%)	7. 5 (3%)			
3. 30 (19%)		3. 13 (8%)	8. 1 (0.6%)			
4. 39 (25%)		4. 12 (8%)	9. 3 (2%)			

Table 19. *Endorsement of Alcohol-Related Problems by T-ACE and ASSIST-3 Items in Females (N = 285).*

T-ACE	Yes N (%)	**ASSIST**	Not at all N (%)	Rarely N (%)	Sometimes N (%)	Almost all the time N (%)
Tolerance	153 (54%)	Loss of control?	208 (73%)	32 (11%)	39 (14%)	6 (2%)
Annoyed?	77 (27%)	Problems?	234 (82%)	22 (8%)	25 (9%)	4 (1%)
Cut down?	148 (52%)	Cravings?	188 (66%)	44 (15%)	41 (14%)	12 (4%)
Eye-opener?	61 (21%)					
Summary Score N (%)		Summary Score N (%)				
0. 72 (25%)		0. 165 (58%)	5. 10 (4%)			
1. 83 (29%)		1. 32 (11%)	6. 13 (5%)			
2. 63 (22%)		2. 29 (10%)	7. 5 (2%)			
3. 38 (13%)		3. 17 (6%)	8. 1 (0.4%)			
4. 29 (10%)		4. 11 (4%)	9. 2 (0.7%)			

Table 20 shows rates of lifetime regular (3+ days/week) use and subsequent responses to ASSIST-3 items for the entire sample. For illicit drug use, rates of lifetime regular (3+ days/week) use were highest for marijuana (60%) and cocaine (35%), followed by heroin (12%), amphetamines (6%), hallucinogens (4%), and inhalants (1%). Similar to alcohol, participants were more likely to report problems related to cravings on the ASSIST-3.

Table 20. *Frequency of Self-Reported Regular Illicit Drug Use and ASSIST Scores by Substance Type (N = 540).*

ASSIST Substance Type	Regular Use* Yes % (N)	Regular Use* No % (N)	ASSIST Question	Not at all % (N)	Rarely % (N)	Sometimes % (N)	Almost all the time % (N)
Marijuana	60% (323)	40% (217)	Loss of control?	75% (242)	11% (37)	11% (36)	2% (8)
			Problems?	89% (288)	7% (22)	3% (10)	0.9% (3)
			Cravings?	54% (175)	14% (46)	19% (61)	13% (41)
Cocaine	35% (190)	65% (350)	Loss of control?	63% (119)	11% (20)	17% (32)	10% (19)
			Problems?	79% (150)	8% (16)	9% (18)	3% (6)
			Cravings?	57% (109)	13% (25)	23% (44)	6% (12)
Heroin	12% (64)	88% (476)	Loss of control?	66% (42)	16% (10)	9% (6)	9% (6)
			Problems?	80% (51)	8% (5)	8% (5)	5% (3)
			Cravings?	55% (35)	11% (7)	14% (9)	20% (13)
Amphetamine	6% (30)	94% (510)	Loss of control?	80% (24)	10% (3)	0% (0)	10% (3)
			Problems?	83% (25)	10% (3)	3% (1)	3% (1)
			Cravings?	77% (23)	3% (1)	13% (4)	7% (2)
Hallucinogen	4% (23)	96% (517)	Loss of control?	96% (22)	0% (0)	0% (0)	4% (1)
			Problems?	91% (21)	4% (1)	0% (0)	4% (1)
			Cravings?	91% (21)	4% (1)	4% (1)	0% (0)
Inhalant	1% (5)	99% (535)	Loss of control?	100% (5)	0% (0)	0% (0)	0% (0)
			Problems?	100% (5)	0% (0)	0% (0)	0% (0)
			Cravings?	100% (5)	0% (0)	0% (0)	0% (0)

*Lifetime history of regular (3+ days/week) use.

Overall, N = 126 (23%) reported marijuana as their primary problem drug, while 14%, 4%, and 1% reported cocaine, heroin, and amphetamine as their primary problem drug,

respectively; N = 56 (10%) self-reported illicit drug use but denied problematic use problematic use of a specific drug. Over half of the sample (N = 285; 53%) self-reported recent (past 30-day) illicit drug use on the CARA. Of those who reported recent illicit drug use, approximately three-fourths (N = 211; 74%) self-reported using one illicit drug. Over one-fifth (11%) self-reported using two different illicit drugs and 2% self-reported using three different illicit drugs (see Table 21). Zero participants self-reported using four illicit drugs and N = 1 self-reported using five illicit drugs. Table 21 displays the response option (i.e., rarely, sometimes, almost all the time) count after summing each ASSIST-3 question among those who endorsed use of multiple illicit drugs, which were used to calculate participants' total ASSIST-3 score. When adding across drugs, cocaine was the major confound for those endorsing use of multiple illicit drugs, and the total (sum) ASSIST-3 scores were largely comprised of ASSIST Q1-Q3 scores for participants' primary drug.

Table 21. *Response Option Count of ASSIST Questions by Primary Substance Type (N = 540).*

Number of Illicit Drugs	% (N)	Substance Type	ASSIST Q1 Score			ASSIST Q2 Score			ASSIST Q3 Score			% (N) Primary Drug
			R1	R2	R3	R1	R2	R3	R1	R2	R3	
2	11% (62)	Marijuana	6	3	0	1	2	0	3	4	5	19% (12)
		Cocaine	10	11	10	6	9	4	6	19	9	60% (37)
		Heroin	3	1	1	2	1	1	2	3	2	11% (7)
		Amphetamine	1	0	1	0	1	0	0	1	1	3% (2)
		Inhalant	0	0	0	0	0	0	0	0	0	0% (0)
		Hallucinogen	0	0	0	0	0	0	0	0	0	0% (0)
		ASSIST Q SUM	20	15	12	9	13	5	11	27	17	--
		Marijuana	0	2	0	0	1	0	0	1	1	18% (2)
		Cocaine	0	1	1	1	0	1	1	1	1	27% (3)
		Heroin	0	1	2	2	0	0	0	1	2	27% (3)

3	2% (11)	Amphetamine	0	0	2	0	1	1	0	2	1	27% (3)
		Inhalant	0	0	0	0	0	0	0	0	0	0% (0)
		Hallucinogen	0	0	1	0	0	0	0	1	0	0% (0)
		ASSIST Q SUM	0	4	5	3	2	2	1	5	5	--

R1 = Rarely; R2 = Sometimes; R3= Almost All the Time

Summarized below are the response rates to problematic illicit drug use as measured by the CAGE-DRUG and ASSIST-3, separately for males (Table 22) and females (Table 23). Similar to alcohol related problems, both males and females were more likely to report problems related to feeling the need to cut down on their drug use on the CAGE-DRUG and experiencing cravings on the ASSIST-3.

Table 22. *Endorsement of Drug-Related Problems by CAGE and ASSIST-3 Items in Males (N = 144).*

CAGE	Yes N (%)	**ASSIST**	Not at all N (%)	Rarely N (%)	Sometimes N (%)	Almost all the time N (%)
Cut down?	66 (46%)	Loss of control?	74 (51%)	20 (14%)	34 (24%)	16 (11%)
Annoyed?	46 (31%)	Problems?	105 (73%)	18 (13%)	17 (12%)	4 (3%)
Guilty?	61 (42%)	Cravings?	48 (33%)	22 (15%)	51 (35%)	23 (16%)
Eye-opener?	65 (45%)					
Summary Score N (%)		Summary Score N (%)				
0. 61 (42%)		0. 40 (28%)	5. 9 (6%)			
1. 8 (6%)		1. 13 (9%)	6. 13 (9%)			
2. 24 (17%)		2. 23 (16%)	7. 7 (5%)			
3. 22 (4%)		3. 12 (8%)	8. 2 (0.4%)			
4. 29 (20%)		4. 23 (16%)	9. 2 (0.4%)			

Table 23. *Endorsement of Drug-Related Problems by CAGE and ASSIST-3 Items in Females (N = 220).*

CAGE	Yes N (%)	**ASSIST**	Not at all N (%)	Rarely N (%)	Sometimes N (%)	Almost all the time N (%)
Cut down?	83 (38%)	Loss of control?	147 (67%)	30 (14%)	29 (13%)	14 (6%)
Annoyed?	61 (28%)	Problems?	179 (81%)	20 (9%)	16 (7%)	5 (2%)
Guilty?	75 (34%)	Cravings?	100 (46%)	32 (15%)	51 (23%)	37 (17%)

Eye-opener? 91 (41%)	
Summary Score N (%)	Summary Score N (%)
0. 95 (43%)	0. 95 (43%) 5. 8 (4%)
1. 33 (15%)	1. 20 (9%) 6. 11 (5%)
2. 34 (16%)	2. 24 (11%) 7. 5 (2%)
3. 23 (11%)	3. 27 (12%) 8. 3 (1%)
4. 35 (16%)	4. 23 (11%) 9. 4 (2%)

Tables 24 – 27 compare the mean ASSIST scores by positive responses to each CAGE, T-ACE, and CAGE-DRUG question. For alcohol, participants scoring positive on each individual CAGE and T-ACE item had significantly higher ASSIST-3 scores compared with participants for whom scored negative on the same individual CAGE and T-ACE item (Table 24 and 25). For males endorsing illicit drug use problems, a similar pattern was found, however, nonsignificant differences were found for participants scoring positive on CAGE-DRUG question one compared to ASSIST question one and for participants scoring positive on CAGE-DRUG questions one and two compared to ASSIST question two (Table 26). For females, nonsignificant differences were found for participants scoring positive on CAGE-DRUG question one compared to ASSIST question two (Table 27).

Table 24. *Comparison of Mean ASSIST-3 Screening Test with Alternative Alcohol CAGE Cut-Off Scores in Men (N=159).*

ASSIST Score Mean	CAGE +1			CAGE +2			CAGE +3			CAGE +4		
	Yes	No	t-value (*p*-value)	Yes	No	t-value (*p*-value)	Yes	No	t-value (*p*-value)	Yes	No	t-value (*p*-value)
ASSIST Q1	0.82	0.15	-4.11 (<0.001)	1.00	0.19	-6.04 (<0.001)	1.12	0.29	-6.19 (<0.001)	1.33	0.43	-5.84 (<0.001)
ASSSIT Q2	0.50	0.13	-2.59 (0.010)	0.61	0.13	-3.94 (<0.001)	0.71	0.17	-4.51 (<0.001)	0.90	0.25	-4.76 (<0.001)
ASSIST Q3	1.04	0.30	-4.06 (<0.001)	1.22	0.38	-5.48 (<0.001)	1.32	0.50	-5.28 (<0.001)	1.54	0.63	-5.03 (<0.001)
ASSIST Sum	2.35	0.58	-4.31 (<0.001)	2.83	0.70	-6.25 (<0.001)	3.14	0.96	-6.44 (<0.001)	3.77	1.30	-6.27 (<0.001)

ASSIST Q1 = loss of control; ASSIST Q2 = problems; ASSIST Q3 = cravings.

Table 25. *Comparison of Mean ASSIST-3 Screening Test with Alternative Alcohol T-ACE Cut-Off Scores in Men (N=285).*

ASSIST Score Mean	TACE +1			TACE +2			TACE +3			TACE +4		
	Yes	No	t-value (*p*-value)	Yes	No	t-value (*p*-value)	Yes	No	t-value (*p*-value)	Yes	No	t-value (*p*-value)
ASSIST Q1*	0.56	0.11	-4.24 (<0.001)	0.82	0.14	-7.91 (<0.001)	1.03	0.27	-7.35 (<0.001)	1.34	0.35	-6.80 (<0.001)
ASSSIT Q2*	0.38	0.06	-3.49 (<0.001)	0.56	0.07	-6.44 (<0.001)	0.78	0.15	-7.13 (<0.001)	1.00	0.21	-6.23 (<0.001)
ASSIST Q3*	0.72	0.11	-5.29 (<0.001)	0.98	0.23	-7.83 (<0.001)	1.27	0.35	-8.20 (<0.001)	1.66	0.45	-7.62 (<0.001)
ASSIST Sum	1.66	0.28	-5.22 (<0.001)	2.36	0.43	-9.04 (<0.001)	3.07	0.77	-9.23 (<0.001)	4.00	1.01	-8.37 (<0.001)

ASSIST Q1 = loss of control; ASSIST Q2 = problems; ASSIST Q3 = cravings.

Table 26. *Comparison of Mean ASSIST-3 Screening Test with Alternative CAGE-DRUG Scores in Males (N=144).*

ASSIST Score Mean	Drug CAGE +1			Drug CAGE +2			Drug CAGE +3			Drug CAGE +4		
	Yes	No	t-value (*p*-value)	Yes	No	t-value (*p*-value)	Yes	No	t-value (*p*-value)	Yes	No	t-value (*p*-value)
ASSIST Q1*	1.08	0.75	-1.80 (0.074)	1.16	0.71	-2.51 (0.013)	1.37	0.71	-3.62 (0.001)	1.55	0.79	-3.47 (0.001)
ASSSIT Q2*	0.52	0.34	-1.28 (0.204)	0.55	0.33	-1.59 (0.111)	0.73	0.29	-3.19 (0.005)	0.86	0.34	-3.21 (0.011)
ASSIST Q3*	1.61	0.97	-3.62 (<0.001)	1.61	1.04	-3.19 (0.002)	1.67	1.16	-2.68 (0.006)	1.86	1.21	-2.92 (0.004)
ASSIST Sum	3.22	2.07	-2.90 (0.004)	3.32	2.09	-3.15 (0.002)	3.76	2.16	-4.01 (<0.001)	4.28	2.34	-4.06 (<0.001)

ASSIST Q1 = loss of control; ASSIST Q2 = problems; ASSIST Q3 = cravings.

Table 27. *Comparison of Mean ASSIST-3 Screening Test with Alternative CAGE-DRUG Scores in Females (N=220).*

ASSIST Score Mean	Drug CAGE +1			Drug CAGE +2			Drug CAGE +3			Drug CAGE +4		
	Yes	No	t-value (*p*-value)	Yes	No	t-value (*p*-value)	Yes	No	t-value (*p*-value)	Yes	No	t-value (*p*-value)
ASSIST Q1*	0.70	0.45	-1.91 (0.051)	0.88	0.38	-3.99 (<0.001)	1.02	0.44	-4.15 (0.001)	1.14	0.49	-3.89 (0.005)
ASSSIT Q2*	0.35	0.24	-1.15 (0.252)	0.47	0.19	-2.96 (0.006)	0.57	0.21	-3.41 (0.006)	0.77	0.22	-4.46 (0.003)
ASSIST Q3*	1.34	0.82	-3.33 (0.001)	1.49	0.84	-4.21 (<0.001)	1.72	0.90	-4.90 (<0.001)	1.83	0.98	-4.11 (<0.001)
ASSIST Sum	2.38	1.52	-2.77 (0.006)	2.84	1.41	-4.66 (<0.001)	3.31	1.54	-5.23 (<0.001)	3.74	1.68	-5.05 (<0.001)

ASSIST Q1 = loss of control; ASSIST Q2 = problems; ASSIST Q3 = cravings.

Table 28 shows the sensitivity, specificity, PPV, NPV, and AUC for the ASSIST-3 assessing for problem alcohol and/or illicit drug use in the past 3-months. In some cases (as shown in Table 24 – 27; Figures 2-5), in which marginally higher values would have resulted in a different cut-point but the AUCs were similar, the cut-point matching the other substances was chosen in order to simplify interpretation of the test. The optimal cut-points on the ASSIST-3 for

Table 28. *Problem Alcohol and Illicit Drug Use by CAHS and CARA.*

	+CAGE % (N)	+ASSIST % (N)	+CAGE/ ASSIST % (N)	Sen.[1]	Spec.[2]	PPV[3]	NPV[4]	AUC[5]	$\chi 2$
Alcohol (Males N = 159)	57% (90)	55% (87)	38% (60)	0.67	0.61	0.69	0.58	0.72	11.59*
Positive on CAGE ≥2 (ASSIST-3 Cut-Point = 1)									
	+T-ACE % (N)	+ASSIST % (N)	+T-ACE/ ASSIST % (N)	Sen.[1]	Spec.[2]	PPV[3]	NPV[4]	AUC[5]	$\chi 2$
Alcohol (Females N = 285)	46% (130)	42% (120)	39% (87)	0.67	0.79	0.73	0.74	0.76	60.40*
Positive on T-ACE ≥2 (ASSIST-3 Cut-Point = 1)									
	+CAGE-DRUG % (N)	+ASSIST % (N)	+CAGE-DRUG/ ASSIST % (N)	Sen.[1]	Spec.[2]	PPV[3]	NPV[4]	AUC[5]	$\chi 2$
Illicit Drug Use (Males N = 144)	58% (83)	63% (91)	43% (62)	0.75	0.53	0.68	0.60	0.65	11.15*
Positive on CAGE-DRUG ≥1 (ASSIST-3 Cut-Point = 2)									
	+CAGE-DRUG % (N)	+ASSIST % (N)	+CAGE-DRUG/ ASSIST % (N)	Sen.[1]	Spec.[2]	PPV[3]	NPV[4]	AUC[5]	$\chi 2$
Illicit Drug Use (Females N = 220)	57% (125)	48% (105)	32% (71)	0.68	0.53	0.57	0.64	0.62	9.55*
Positive on CAGE-DRUG ≥1 (ASSIST-3 Cut-Point = 2)									

[1]Sen. = Sensitivity; [2]Spec. = Specificity; [3]PPV = Positive Predictive Value; [4]NPV = Negative Predictive Value; AUC = [5]Area Under the Curve

*Denotes a statistically significant $\chi 2$ value ($p < .05$).

detecting “problem use” for alcohol was 1 for both males (Figure 2) and females (Figure 3), while the cut- point on the ASSIST for detecting “problem use” for illicit drug use was 2 for males (Figure 3) and females (Figure 4). The AUC demonstrated acceptable discrimination between the CAGE and ASSIST-3 (0.72) and between the T-ACE and ASSIST-3 (0.76) but demonstrated poor discrimination between the CAGE-DRUG and ASSIST-3 for both males (0.65) and females (0.62). All AUC values were statistically significant ($p < .05$).

For males, when the CAGE was considered the reference standard, sensitivity and specificity was 0.67 and 0.79 for detecting “problem use” for alcohol ASSIST-3 thresholds. The PPV was 0.69 and NPV was only slightly greater than 50% chance classification (0.58). For females, when the T-ACE was considered the reference standard, sensitivity was 0.67 with higher specificity (0.79). PPV (0.73) and NPV (0.74) were similar. When the CAGE-DRUG was considered the reference standard for detecting “problem use” for illicit drug use, ASSIST-3 sensitivity and specificity was 0.75 and 0.53 for males and 0.68 and 0.53 for females. PPV and NPV for males were 0.68 and 0.60, respectively, while PPV and NPV for females were 0.57 and 0.64, respectively.

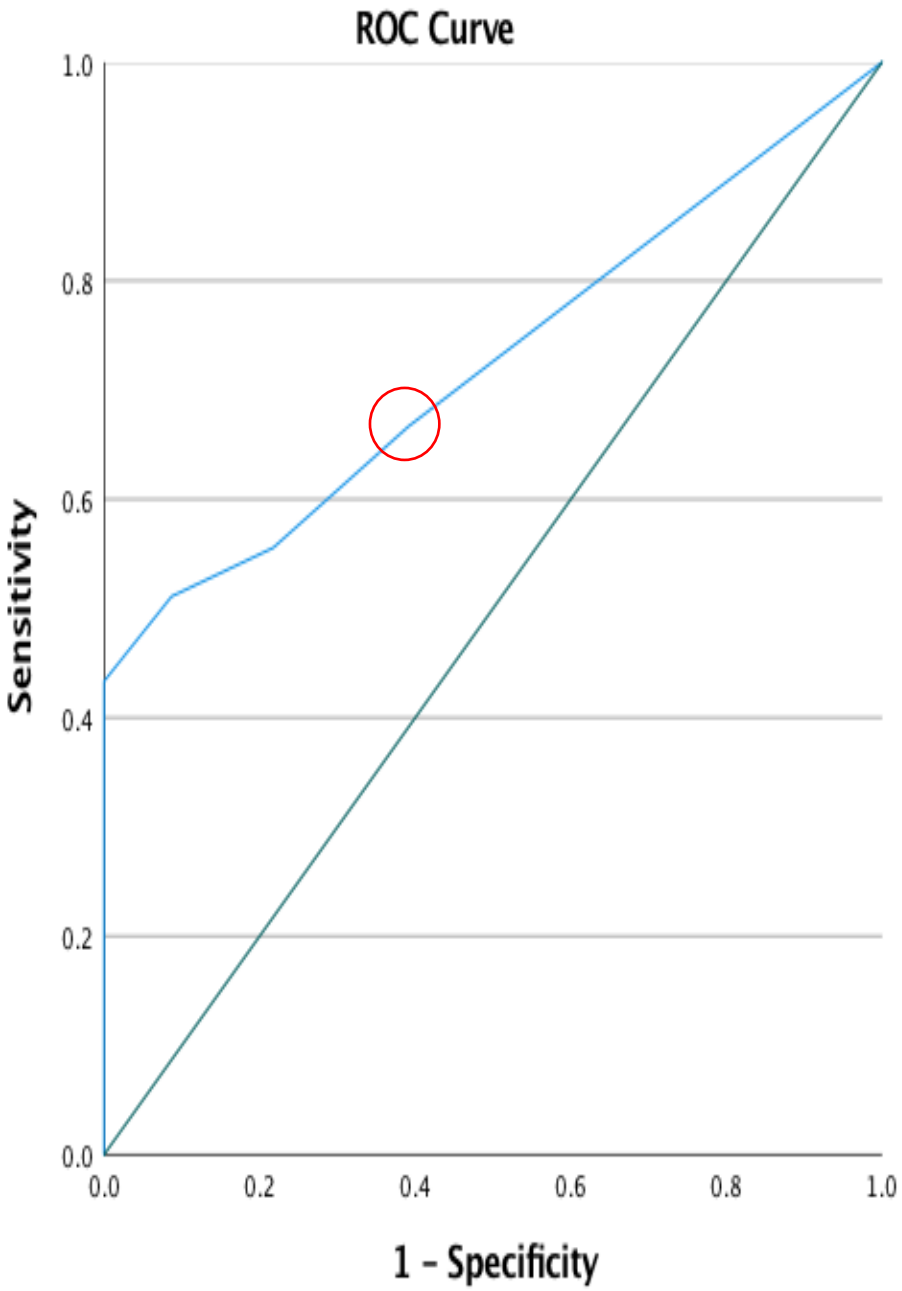

Figure 2. ROC curve for CAGE+ predicting ASSIST-3 alcohol-related problems in males. Optimal cutoff score of 1 is indicated by a circle. The diagonal line indicates the theoretical ROC curve for AUC = .50.

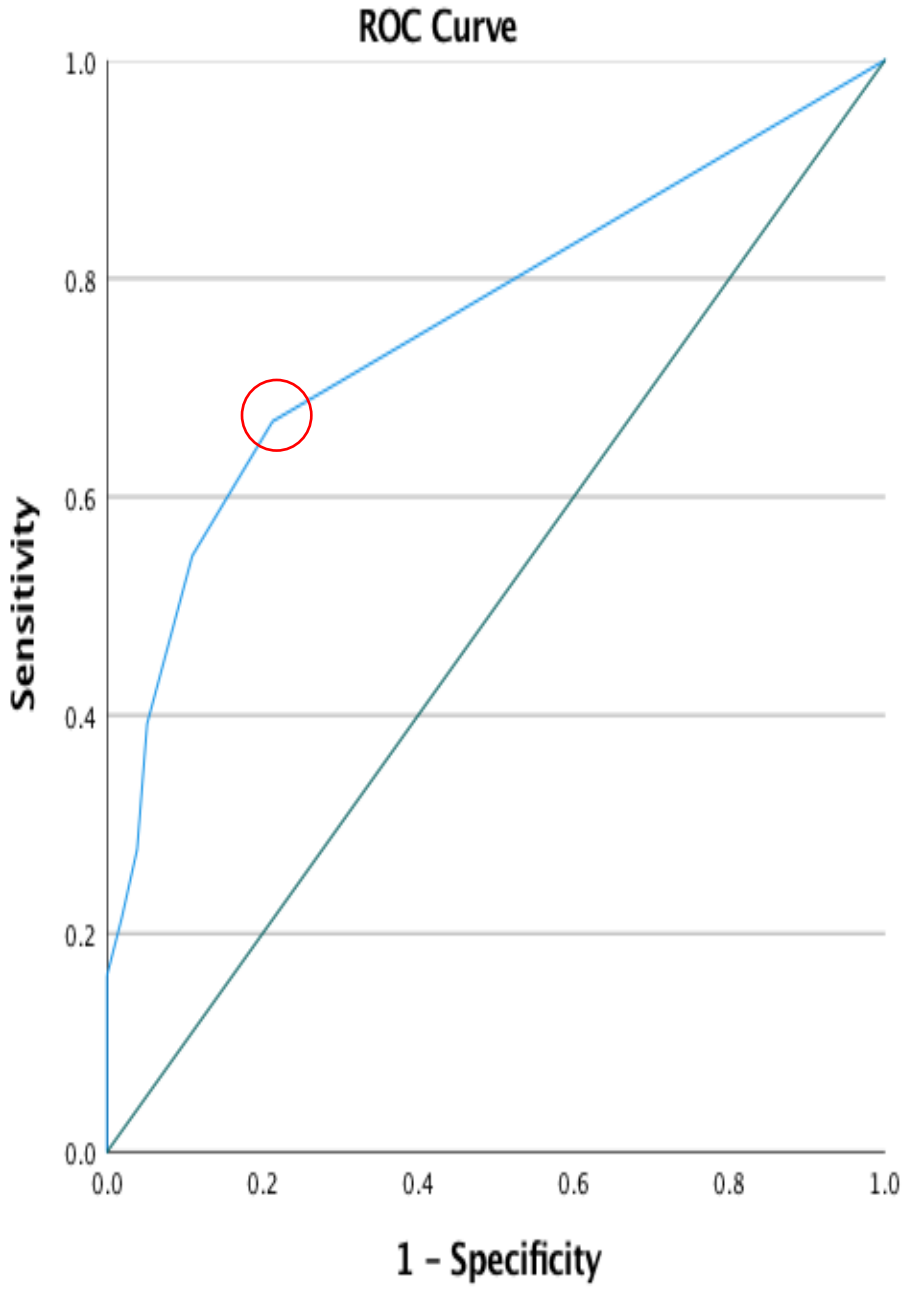

Figure 3. ROC curve for T-ACE+ predicting ASSIST-3 alcohol-related problems in females. Optimal cutoff score of 1 is indicated by a circle. The diagonal line indicates the theoretical ROC curve for AUC = .50.

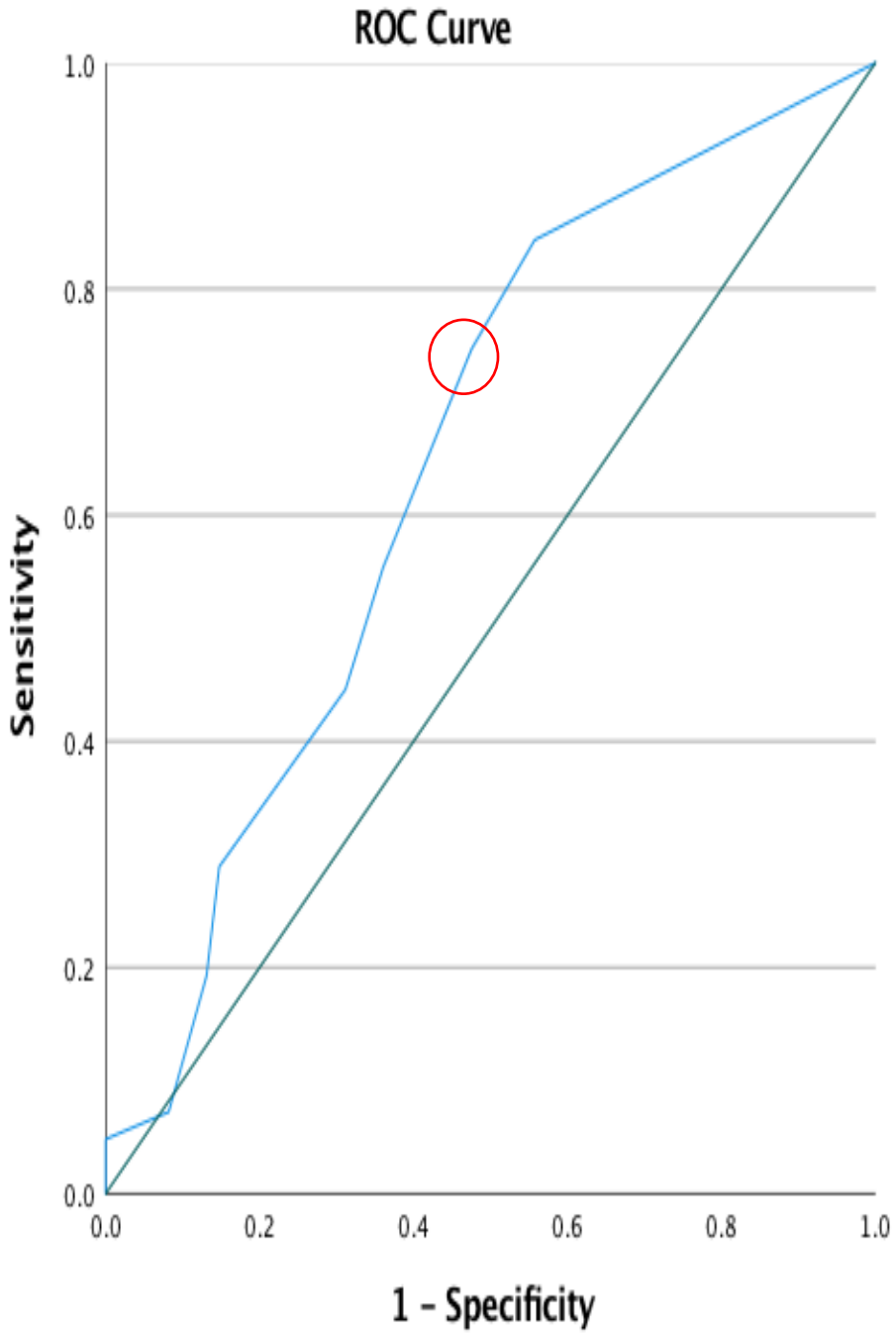

Figure 4. ROC curve for CAGE-DRUG+ predicting ASSIST-3 illicit drug-related problems in males. Optimal cutoff score of 2 is indicated by a circle. The diagonal line indicates the theoretical ROC curve for AUC = .50.

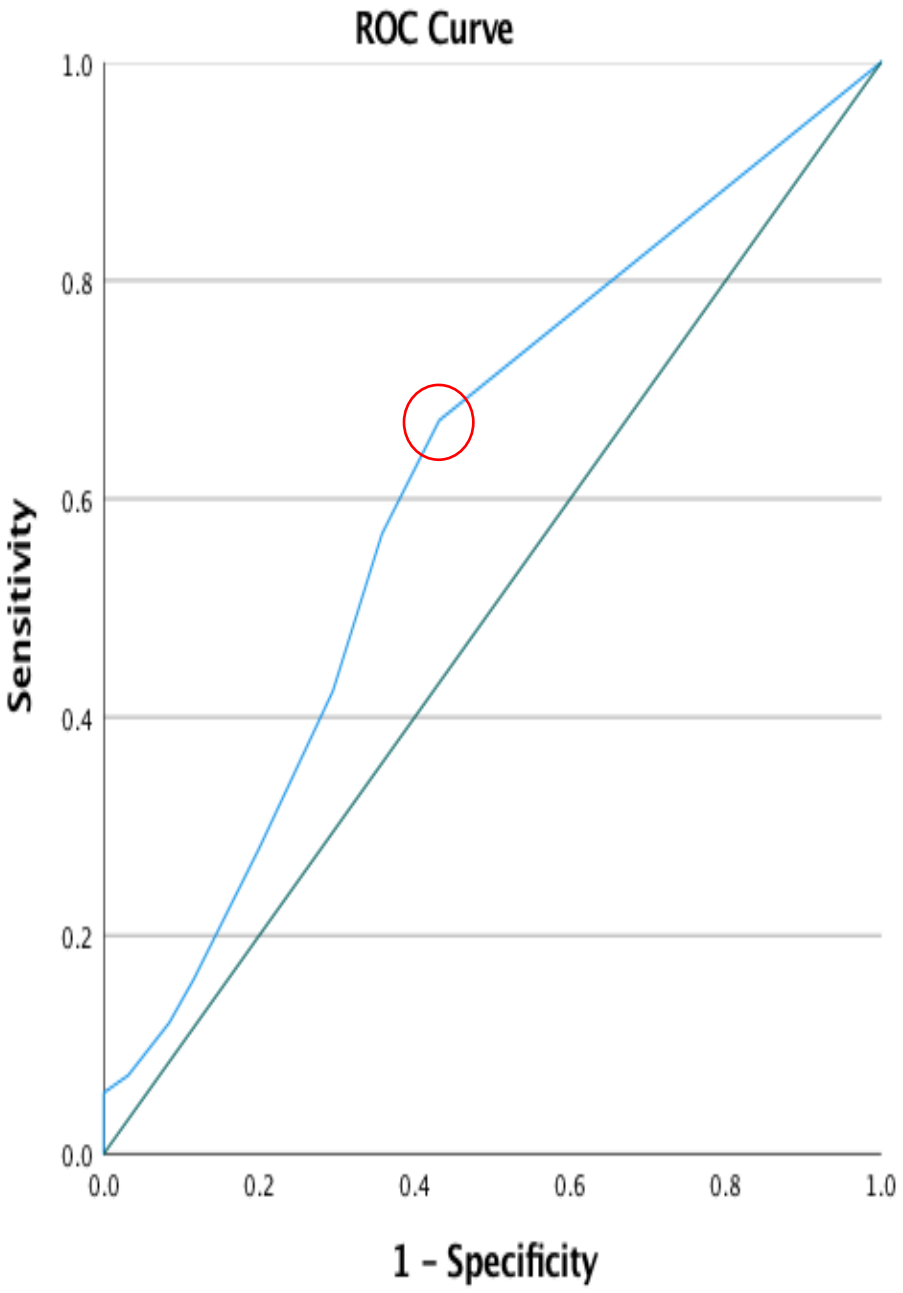

Figure 5. ROC curve for CAGE-DRUG+ predicting ASSIST-3 illicit drug-related problems in females. Optimal cutoff score of 2 is indicated by a circle. The diagonal line indicates the theoretical ROC curve for AUC = .50.

Hypothesis 1. Participants will be more likely to score positive for problematic substance use on the anonymous, computer-administered screener (i.e., CAGE, T-ACE, CAGE-DRUG) compared to the confidential, computerized assessment (i.e., ASSIST-3). Participants were more likely to score positive on the CAGE compared to the ASSIST-3 for both males ($\chi 2$ (1, N = 159) = 11.59 p = 0.001) and females ($\chi 2$ (1, N = 285) = 60.40 p < 0.001), supporting the hypothesized associations. Males were more likely to score positive for illicit drug-related problems on the ASSIST-3 compared to CAGE-DRUG ($\chi 2$ (1, N = 144) = 11.15 p = 0.001), which does not support the hypothesized association. However, females were more likely to score positive for illicit drug-related problems on the CAGE-DRUG compared to the ASSIST-3 ($\chi 2$ (1, N = 220) = 9.55 p < 0.001), which supports the hypothesized association.

Aim 3: Agreement Between Self-Reports of Recent (Past 30-Day) Drug Use and Non-medical Prescription Drug Use by Urinalysis and IARA.

For the present analyses, positive tests for opiates with self-reported heroin use (i.e., the heroin metabolite, 6-Monoacetylmorphine) absent of self-reported non-medical prescription opiate use were classified as heroin-positive (N=54); tests positive for other opioids (codeine, oxycontin, oxycodone, hydrocodone) absent of self-reported heroin use were classified as positive for non-medical prescription opioid use (N = 83). Also, positive tests for amphetamines and methamphetamines (N=8) were classified as amphetamine positive.

Table 29 shows rates of recent (past 30-days) substance use by personal interview (IARA) and UDS. Marijuana was most prevalent, with approximately one-third (N = 154) of participants self-reporting recreational use, followed by cocaine use (14%; N = 64). Positive UDS rates were 30% for marijuana and 20% for cocaine, with 89% overall concordance rate (i.e., agreement between self-reported use on the TLFB and UDS results) for cocaine use and 85% for marijuana use. Both *k* and cond. *k* values indicate acceptable concordance. IARA

Table 29. *Concordance of TLFB Self-Reported Illicit & Non-medical Prescription Drug Use to UDS results.*

	+IARA % (N)	+UDS % (N)	Concor-dance Positive % (N)	Concor-dance Negative % (N)	Overall Concor-dance % (N)	Sens.[1]	Spec.[2]	PPV[3]	NPV[4]
Marijuana (N = 465) k=0.66 cond. k=0.62	33% (154)	30% (141)	24% (113)	61% (283)	85% (396)	.80	.87	.73	.91
Heroin / Opiates (N = 405) k=0.25 cond. k=0.22	24% (97)	20% (83)	9% (37)	65% (262)	74% (299)	.45	.81	.38	.85
Opiates (N = 405) k=0.12 cond. k=0.12	20% (79)	20% (83)	6% (24)	66% (268)	72% (292)	.29	.83	.30	.82
Cocaine (N=465) k=0.59 cond. k=0.49	14% (64)	20% (91)	11% (51)	78% (361)	89% (412)	.56	.97	.80	.90
Heroin (N = 405) k=0.32 cond. k=0.21	4% (18)	13% (54)	3% (13)	86% (347)	89% (360)	.24	.99	.72	.89
Benzodiazepines (N = 405) k=0.19 cond. k=0.12	4% (17)	16% (65)	3% (10)	81% (330)	84% (340)	.15	.98	.59	.86
Amphetamines (N = 405) k=0.19 cond. k=0.12	0.5% (2)	2% (8)	0.2% (1)	97% (393)	97% (394)	.13	1.00	.50	.98

[1]Sen. = Sensitivity
[2]Spec. = Specificity
[3]PPV = Positive Predictive Value
[4]NPV = Negative Predictive Value

sensitivity for detecting marijuana was 0.80 and 0.56 for cocaine. The specificity of was high for each drug, ranging from 0.87 for marijuana to 0.97 for cocaine. The Positive Predictive Value

(PPV) was high for cocaine (0.80) and slightly lower for marijuana (0.73), and Negative Predictive Values (NPV) were high for each drug ranging from 0.90 for cocaine to 0.91 for marijuana.

For the N = 405 with substance specific non-medical prescription drug use information, including urine assay-opiate results, approximately one-fourth (N = 97) self-reported either using heroin or non-medical opioid medication, while less than 5% (N = 18) self-reported only heroin use. Rates of reported recent non-medical benzodiazepine and non-medical amphetamine use were 4% and 0.5%, respectively. Positive UDS rates were 20% for heroin / opiates and opiates, 13% for heroin, 16% for benzodiazepines, and 2% for non-medical amphetamine use.

Overall concordance rates for each substance type ranged from 97% for non-medical amphetamine use, 89% for heroin use, 84% for non-medical benzodiazepine use, 74% for heroin / opiate use, and to 72% for non-medical opiate drug use. However, the low value of both *k* and cond. *k* indicates that positive concordance was poor (see Table 29). IARA sensitivity for detecting heroin / opiates was 0.45, 0.29 for non-medical opiate use, 0.24 for heroin, 0.15 for non-medical benzodiazepine use, and 0.13 for non-medical amphetamine use. The specificity of was high for each drug, ranging from 0.81 for heroin / opiates to 1.00 for amphetamines. The PPV for opiates (0.30) and heroin / opiates (0.38) was moderate, with higher PPVs for both amphetamines (0.50) and benzodiazepines (0.59), and even higher PPV for opiates (0.72). Negative Predictive Values (NPV) were high for each drug ranging from 0.82 for opiates to 0.98 for non-medical amphetamine use.

For the sample based on illicit drugs, urinalysis yielded higher rates of cocaine use than self-report; however, the TLFB yielded higher rates of marijuana use compared to urinalysis. For the sample based on prescription drug use, urinalysis generally yielder higher rates than self-

report. The only case in which self-reported use was higher than that detected by urinalysis was for reports of heroine / opiate use.

In order to examine discordant results on the TLFB and the UDS, individuals were considered positive for each substance if they either reported use on TLFB or had a positive OFT result. Table 30 displays the discordance between the TLFB and UDS across each substance.

Table 30. *Discordance Between TLFB and UDS results for marijuana (N = 182), heroin / opiates (N = 143), opiates (N = 138), cocaine (N = 104), benzodiazepines (N = 72), heroin (N = 59), and amphetamines (N = 9).*

	+IARA -UDS % (N)	-IARA +UDS % (N)	Total Discordance % (N)
Marijuana	23% (41)	15% (28)	38% (69)
Heroin / Opiates	42% (60)	32% (46)	74% (106)
Opiates	40% (55)	43% (59)	82% (113)
Cocaine	13% (13)	38% (40)	51% (53)
Benzodiazepines	10% (7)	42% (30)	51% (37)
Heroin	8% (5)	69% (41)	78% (46)
Amphetamines	11% (1)	78% (7)	89% (8)

Among the subsample of participants with any positive marijuana (N = 182), heroin / opiate (N = 143), opiate (N = 138), cocaine (N = 104), benzodiazepine (N = 72), heroin (N = 59), and amphetamine (N = 9) use, total discordance rates between the two measures were 38% for

marijuana, 74% for heroin / opiates, 82% for opiates, 51% for cocaine and benzodiazepines, 78% for heroin, and 89% for non-medical amphetamine use.

Hypothesis 1*:* Proportion of illicit drug use and non-medical prescription drug use will be higher by urine assay compared to self-report measure. Rates of cocaine ($\chi 2$ (1, N = 465) = 170.41, $p < 0.001$), heroin ($\chi 2$ (1, N = 405) = 56.71, $p < 0.001$), non-medical benzodiazepine ($\chi 2$ (1, N = 405) = 23.93, $p < 0.001$), non-medical amphetamine ($\chi 2$ (1, N = 405) = 23.75 $p < 0.001$), and opiate ($\chi 2$ (1, N = 405) = 5.95, $p = 0.015$) use were higher on the UDS compared to IARA, supporting the hypothesized association. Rates of marijuana ($\chi 2$ (1, N = 465) = 202.02, $p < 0.001$) and heroin / opiates ($\chi 2$ (1, N = 405) = 23.80, $p < 0.001$) use were higher on the IARA compared to UDS results. These results do support the hypothesized associations.

Aim 4: Multivariable Analysis

All of the independent variables included in the multivariate analysis to identify correlates of any recent (past 30-day) alcohol and other drug use are summarized in Table 31.

Table 31. *Variables Included in Multivariable Logistic Regression.*

Demographics	**Medical**	
Age[3,5,6,7,8,9,10,11,12,15,16,17,19]	Anxiety[1,2,3,4,5,6,7,8,9,10,11,12,13,14,18,19,20]	
Gender[2,7,9,12,13,14,16,20]	Depression[1,3,4,6,7,8,9,10,11,12,14,16,17,18,19,20]	
Race[1,3,4,6,9,10,11,18,19]	Psychological Distress (K10)[1,4,6,7,8,9,9,10,11,13,14,15,16,17,19,20]	
Ethnicity[13,14]	2+ Medical Conditions[2,5,9,10,11,12,13,14,18,19,20]	
Education[2,4,5,9,10,11,12,13,14,15,16,17,18]	COPD	High Blood Pressure
Insurance[2,4,5,7,8,11,13,14,16]	Liver Disease	Pancreatitis
Employment[3,5,7,8,16]	Hepatitis	High Cholesterol
Marital Status[1,4,7, 10,14]	Heart Disease	Chronic Pain
Quality of Sleep[6,9,15,17]	Migraines	Asthma
Sleep Problems[1,2,4,7,8,9,10,11,13,14,16,18,19,20]	Arthritis	
Taking Sleep Aids[4,7,9,10,11,14,16,18,19,20]		

Variables selected for any recent alcohol use: CAHS[1], CARA[2], IARA[3]
Variables selected for any recent binge drinking: CAHS[4], IARA[5]
Variables selected for any recent illicit drug use: CAHS[6], CARA[7], IARA[8]
Variables selected for any recent illicit drug use: CAHS[9], CARA[10], IARA[11]
Variables selected for mean number of drinks per week: CAHS[12], CARA[13], IARA[14]
Variables selected mean days of illicit drug use per week: CAHS[15], CARA[16], IARA[17]
Variables selected mean days of prescription drug misuse per week: CAHS[18], CARA[19], IARA[20]

Based on a classification threshold predicted probability of target group membership of 0.5, models for any recent (past 30-day) alcohol and binge drinking by assessment measure were statistically significant. The Nagelkerke pseudo R^2 suggested that the model(s) accounted between approximately 5% and 6% of the total variance in any recent (past 30-day) alcohol and between 5% and 7% in any recent binge drinking (see Table 32).

Classification success for the cases based on a classification cutoff value of 0.5 for predicting any recent (past 30-day alcohol use) was moderately high for the CAHS and IARA, with overall prediction success rates of 88% and 80%, respectively; prediction success rate for the CARA was less at 63%. For any recent (past 30-day) binge drinking, the CAHS correctly classified 74% of participants while the IARA only correctly classified 56%.

As shown in Table 32, significant predictors of any recent (past 30-day) alcohol use were taking sleep aids in the past 30 days and being from a racial minority group for the CAHS; having government insurance and being female were significant predictors for the CARA, while endorsing symptoms of depression and being unemployed were significant predictors for the IARA. Having >12 years of education and being from a racial minority group were associated with any recent (past 30-day) binge drinking on the CAHS, while being older age, having government insurance, and having >12 years of education were all associated with any recent (past 30-day) binge drinking on the IARA. No other predictors exerted a unique effect on any recent (past 30-day) alcohol use or binge drinking (all *p*s >.05).

Table 33 shows that, based on a classification threshold predicted probability of target group membership of 0.5, models for any recent (past 30-day) illicit drug use and prescription drug misuse by assessment measure were statistically significant. The Nagelkerke pseudo R^2 suggested that the model(s) accounted between approximately 8% and 16% of the total variance

in any recent (past 30-day) illicit drug use and prescription drug misuse. Classification success for the cases based on a classification cutoff value of 0.5 for predicting any recent (past 30-day) illicit drug use was moderate, with an overall prediction success rate of 66% for the CAHS, 61% for the CARA, and 59% for the IARA. Overall classification prediction success rate was higher for any recent (past 30-day) prescription drug misuse with 76%, 74%, and 73% rates for the CAHS, CARA, and IARA, respectively.

Significant predictors for any recent (past 30-day) illicit drug use were endorsing moderate-severe psychological distress and being older age on the CAHS, while trouble staying asleep in the past 30 days, being older and female, and endorsing moderate-severe psychological distress were associated with any recent (past 30-day) illicit drug use on the CARA. Trouble falling asleep in the past 30 days, being older, and endorsing moderate-severe psychological distress were associated with any recent illicit drug use on the IARA (see Table 33). Significant predictors for any recent (past 30-day) prescription drug misuse were taking sleep aids, having >12 years of education, being older age, and having 2+ medical conditions on the CAHS. Trouble falling asleep in the past 30 days, having 2+ medical conditions, being married, and being older age were all associated with any recent past prescription drug misuse on the CARA, while trouble falling asleep in the past 30 days and endorsing symptoms of anxiety were significant predictors on the IARA. No other predictors exerted a unique effect on any recent (past 30-day) illicit drug or prescription drug misuse (all *p*s > .05) (see Table 33).

All of the independent variables summarized in Table 31 were also used to identify correlates of alcohol QF and frequency of other drug use per week in the past 30 days. The overall model(s) significantly predicted both alcohol and other drug use per week during the past 30 days. Together, these predictors accounted between 7% and 8% of the variance in alcohol use

per week (see Table 34) during the past 30 days, between 8% and 11% of the variance in illicit drug use per week, and between 5% and 10% in prescription drug misuse (see Table 35).

Being female and endorsing symptoms of anxiety were significant predictors of alcohol use per week during the past 30 days on the CAHS, while being female, >12 years of education, and having government insurance were significant predictors of alcohol use per week during the past 30 days on the CARA. Trouble staying asleep in the past 30 days, being female, endorsing symptoms of anxiety, having >12 of education, and being married were significant predictors of alcohol use per week on the IARA (see Table 34).

As shown in Table 35, being older, having >12 years of education, and endorsing moderate-severe psychological distress were associated with illicit drug use per week during the past 30 days on the CAHS; trouble staying asleep in the past 30 days, being older and female, having >12 years of education, and endorsing moderate-severe psychological distress were significant predictors on the CARA, while being older and having >12 years of education were significant predictors on the IARA. Taking sleep aids in the past 30 days and having 2+ medical conditions were associated with prescription drug misuse per week on the CAHS, while trouble falling asleep and taking sleep in the past 30 days, being older, having 2+ medical conditions, and endorsing moderate-severe psychological distress were significant predictors on the CARA. Taking sleep aids in the past 30 days and endorsing symptoms of anxiety were significant predictors of prescription drug misuse per week on the IARA. No other predictors exerted a unique effect on days of alcohol or other drug use per week during the past 30 days (all *p*s > .05).

Table 32. *Multivariable Logistic Regression by Assessment Measure for Alcohol (N=540).*

Substance Use Variable	Assessment Measure	Covariate	B	Standard Error	*p*-value	95% CI for Odds Ratio Lower	Exp(B)	Upper	Nagelkerke R^2	Model χ^2
Any Recent (Past 30-Day) Alcohol Use	CAHS	Sleep Aids	0.73	0.32	**0.024**	1.10	2.07	3.89	0.05	10.29*
		Race	-0.69	0.34	**0.039**	0.26	0.50	0.97		
	CARA	Sleep Problems	0.45	0.25	0.069	0.97	1.57	2.54		
		Insurance	-0.57	0.27	**0.031**	0.34	0.56	0.95	0.06	16.94*
		Gender	-0.45	0.23	**0.047**	0.41	0.64	1.00		
		Education	-0.46	0.24	0.054	0.40	0.63	1.01		
	IARA	Depression	0.58	0.25	**0.021**	1.09	1.79	2.94	0.06	15.06*
		Employment	0.85	0.35	**0.015**	1.18	2.35	4.67		
Any Recent (Past 30-Day) Binge Drinking	CAHS	Race	-0.76	0.30	**0.011**	0.26	0.47	0.84	0.07	18.06*
		Education	0.81	0.25	**0.001**	1.39	2.25	3.66		
		Age	-0.02	0.01	**0.046**	0.96	0.98	1.00		
	IARA	Insurance	0.59	0.27	**0.029**	1.06	1.80	3.06	0.05	13.24*
		Education	0.50	0.23	**0.034**	1.04	1.64	2.60		

*Denotes a statistically significant χ^2 ($p < .05$).

Table 33. *Multivariable Logistic Regression by Assessment Measure for Other Drugs (N=540).*

Substance Use Variable	Assessment Measure	Covariate	B	Standard Error	*p*-value	95% CI for Odds Ratio Lower	Exp(B)	Upper	Nagelkerke R^2	Model χ^2
Any Recent (Past 30-Day) Illicit Drug Use	CAHS	Sleep Quality	0.35	0.19	0.071	0.97	1.41	2.05	0.09	*34.93
		K10	-0.62	0.20	**0.002**	0.37	0.54	0.80		
		Age	-0.03	0.01	**<0.001**	0.95	0.97	0.99		
	CARA	Sleep Problems	-0.58	0.24	**0.016**	0.35	0.56	0.90	0.10	*29.50
		Gender	0.65	0.23	**0.004**	1.23	1.91	2.97		
		K10	-0.64	0.24	**0.009**	0.33	0.54	0.86		
		Age	-0.03	0.01	**0.006**	0.95	0.97	0.99		
	IARA	Sleep Problems	-0.48	0.25	0.054	0.38	0.62	1.01	0.08	*23.83
		K10	-0.58	0.24	**0.014**	0.35	0.56	0.89		
		Age	-0.03	0.01	**0.006**	0.95	0.97	0.99		
Any Recent (Past 30-Day) Prescription Drug Misuse	CAHS	Sleep Quality	0.43	0.24	0.074	0.96	1.54	2.47	0.14	*40.97
		Sleep Aids	-0.75	0.25	**0.002**	0.29	0.47	0.76		
		Education	-0.54	0.25	**0.030**	0.36	0.59	0.95		
		2+ Medical	-1.07	0.29	**<0.001**	0.19	0.34	0.61		
		Age	-0.03	0.01	**0.026**	0.95	0.98	1.00		
	CARA	Sleep Problems	-0.78	0.31	**0.012**	0.25	0.46	0.84	0.16	*47.19
		Anxiety	-0.46	0.26	0.074	0.38	0.63	1.05		
		2+ Medical	-0.62	0.28	**0.023**	0.31	0.54	0.92		
		Marital	-1.06	0.35	**0.002**	0.18	0.35	0.69		
		Age	-0.04	0.01	**0.002**	0.95	0.97	0.99		
		K10	-0.47	0.26	0.070	0.37	0.62	1.04		
	IARA	Sleep Problems	-1.11	0.34	**0.001**	0.17	0.33	0.64	0.11	*29.77
		Anxiety	-0.82	0.25	**0.001**	0.27	0.44	0.72		
		Insurance	0.54	0.33	0.102	0.90	1.71	3.25		

*Denotes a statistically significant χ^2 ($p < .05$).

Table 34. *Multivariable Linear Regression by Assessment Measure for Alcohol (N=540).*

Substance Use Variable	Assessment Measure	Covariate	B	SE B	β	t	*p*-value	95% CI for B Lower	95% CI for B Upper	F	*p*-value	R^2
	CAHS	Gender	-0.65	0.17	-0.19	-3.91	**<0.001**	-0.98	-0.32	5.89	**<0.001**	0.08
		Anxiety	-0.53	0.17	-0.15	-3.16	**0.002**	-0.86	-0.20			
		Sleep Problems	0.40	0.24	0.09	1.71	0.089	-0.06	0.87			
	CARA	Gender	-0.58	0.21	-0.14	-2.76	**0.006**	-1.00	-0.17	6.51	**<0.001**	0.07
*Alcohol		Education	-0.62	0.23	-0.14	-2.71	**0.007**	-1.07	-0.17			
		Insurance	-0.59	0.26	-0.12	-2.31	**0.022**	-1.09	-0.09			
		Sleep Problems	0.72	0.36	0.11	2.00	**0.046**	0.01	1.42			
		Gender	-0.82	0.32	-0.13	-2.60	**0.010**	-1.44	-0.20			
	IARA	Anxiety	-0.75	0.32	-0.12	-2.30	**0.022**	-1.39	-0.11	5.81	**<0.001**	0.08
		Marital	1.21	0.51	0.12	2.36	**0.019**	0.20	2.22			
		Education	-0.80	0.34	-0.12	-2.33	**0.020**	-1.47	-0.13			

***Mean Days of Alcohol Use per Week During the Past 30 Days*

Table 35. *Multivariable Linear Regression by Assessment Measure for Other Drugs (N=540).*

Substance Use Variable	Assessment Measure	Covariate	B	SE B	β	t	*p*-value	95% CI for B Lower	95% CI for B Upper	F	*p*-value	R^2
*Illicit Drug	CAHS	Age	-0.06	0.01	-0.28	-5.84	**<0.001**	-0.08	-0.04	14.64	**<0.001**	0.10
		Education	-0.55	0.25	-0.11	-2.24	**0.025**	-1.03	-0.07			
		K10	0.47	0.25	0.09	1.91	**0.056**	-0.01	0.96			
	CARA	Sleep Problems	0.29	0.14	0.10	1.99	**0.047**	0.00	0.57	6.92	**<0.001**	0.09
		Age	-0.02	0.01	-0.20	-3.99	**<0.001**	-0.04	-0.01			
		Gender	-0.35	0.13	-0.14	-2.74	**0.006**	-0.60	0.10			
		Education	-0.27	0.14	-0.10	-2.00	**0.046**	-0.54	-0.01			
		K10	0.26	0.14	0.10	1.90	**0.058**	-0.01	0.52			
	IARA	Sleep Quality	-0.42	0.24	-0.08	-1.75	0.080	-0.89	0.05	12.36	**<0.001**	0.08
		Education	-0.55	0.25	-0.11	-2.20	**0.028**	-1.04	-0.06			
		Age	-0.06	0.01	-0.27	-5.55	**<0.001**	-0.08	-0.04			
**Prescription Drug	CAHS	Sleep Aids + Medical	0.37	0.13	0.14	2.87	**0.004**	0.12	0.62	5.74	**<0.001**	0.05
		Race	0.30	0.13	0.11	2.25	**0.025**	0.04	0.56			
		Education	-0.27	0.16	-0.08	-1.68	0.094	-0.59	0.05			
		Sleep Problems	0.24	0.14	0.09	1.78	0.076	-0.03	0.51	9.70	**<0.001**	0.10
	CARA	Sleep Aids	0.25	0.12	0.09	2.09	**0.037**	0.01	0.48			
		Age	0.25	0.10	0.11	2.44	**0.015**	0.05	0.46			
		2+ Medical	-0.01	0.01	-0.10	-2.23	**0.026**	-0.02	-0.00			
		Race	0.40	0.11	0.16	3.54	**<0.001**	0.18	0.62			
		K10	-0.24	0.13	-0.08	-1.80	0.072	-0.49	0.02			
			0.32	0.11	0.12	2.79	**0.005**	0.09	0.54			
	IARA	Sleep Aids	0.38	0.15	0.11	2.51	**0.012**	0.08	0.68	13.09	**<0.001**	0.05
		Anxiety	0.55	0.16	0.16	3.52	**<0.001**	0.24	0.85			

*****Mean *Days of Illicit Drug Use per Week During the Past 30 Days*

***Mean Days of Prescription Drug Misuse per Week During the Past 30 Days*

Discussion

From a public health perspective, primary care settings offer an excellent opportunity to identify and address patient substance use and related problems. For large scale implementation, however, efficient tools are needed that are both valid and reliable. The present study compared computerized anonymous and confidential surveys to a confidential, interviewer-administered assessment in a primary care sample of heavy/problem alcohol and/or drug using males and females. This study also compared self-reports of recent substance use to urine drug assay findings. Finally, this study identified correlates of any recent (past 30-day) use as well as correlates of recent alcohol quantity-frequency and other drug use frequency per week in the past 30 days.

Any recent (past 30-day) alcohol/drug use (i.e., illicit drug use and/or prescription drug misuse) as well as recent alcohol quantity-frequency and other drug use frequency were measured by all three methods of administration. The computerized, anonymous survey and interviewer-administered assessment also measured frequency of recent binge drinking, while the computerized, anonymous survey and computerized, confidential survey measured problem alcohol and/or drug use.

Overall, prevalence of any substance use were higher by anonymous computer survey than those found by the other two methods of administration. Specifically, any recent (past 30-day) alcohol use was highest on the computerized, anonymous survey while rates of any recent (past 30-day) illicit drug use were highest on the confidential, computerized assessment. Rates of binge drinking were significantly higher on the computerized, anonymous survey compared to interviewer-administered assessment; and rates of alcohol and/or drug use related problems were significantly higher on the anonymous computerized survey versus confidential computerized

survey. Quantity-frequency of alcohol use and frequency of prescription drug misuse per week were significantly higher on the interviewer-administered assessment compared to the other two methods of administration. While frequency of illicit drug use per week was highest on the interviewer-administered assessment, significant differences were only found when compared to the computerized, confidential survey.

Sample Representativeness

It is important to recognize that the present study looked only at individuals who qualified for a RCT targeting heavy/problem alcohol and/or illicit drug use or prescription medication misuse. Primary care patients who were not at risk are absent from the sample. Further, of those who met RCT criteria, only 53% gave informed consent and participated in the clinical trial. Previous research by members of our research team found RCT consenters differed from non-consenters on a variety of measures. Specifically, RCT participants were more likely to endorse prescription drug misuse and problems related to drug use. They also endorsed a greater number and more severe psychosocial and mental health problems (Kelpin et al., 2018). As a result, the representativeness of the present study sample and generalizability of present study findings to broader patient groups (e.g., all primary care patients) must be made with caution.

Anonymous Survey vs. Personal Interview (CAHS and IARA)

Any Recent Alcohol Use and Binge Drinking

For alcohol, participants were more likely to report recent alcohol use by CAHS than by IARA (i.e., the TLFB). For any recent use, the difference was modest (86% by CAHS and 79% by IARA), as was the difference in number of days of recent alcohol use. For binge drinking, however, the difference was more striking, with nearly twice as many participants reporting any recent binge drinking by anonymous survey (71% on CAHS) as compared to personal interview

(43% by IARA). This difference was also seen in the number of recent binge drinking days reported, with a mean of 5.37 days on the CAHS compared to 3.52 days on the IARA.

A similar pattern was reported by Gryczynski, et al (2017) in a multi-site study of adult primary care patients. They found an anonymous tablet-based alcohol screen yielded higher self-reported alcohol use than a subsequent interviewer-administered version of the same questions (48% vs. 37%; $p < 0.001$). Interestingly, this pattern was not seen when the interviewer mode of administration preceded the computer-based survey (e.g., all *p*s > 0.05). For the present study, the anonymous, computerized survey was always administered prior to the personal interview, as it was the screener that determined RCT eligibility.

Similarly, in a research study with postpartum women, Beatty et al., (2014) found that those randomly assigned to an anonymous consent condition disclosed significantly more alcohol and drug use than those completing traditional confidential consent procedures. Underreporting of alcohol and drug use is of particular concern in pregnant women, where fear of adverse societal or legal consequences is often elevated (Grekin et al., 2010). The current study suggests similar under-reporting of heavy/binge may not be limited to postpartum women.

In the field of alcohol research, the TLFB has long been considered the most reliable and valid method of assessment for quantity and frequency of alcohol use (Sobell & Sobell, 2003; Sobell & Sobell, 2008; Sobell et al., 2001; Sobell et al., 2003). The TLFB is rarely used in routine clinical practice (e.g., as a screening tool) as it is too time consuming and requires more staff training than other measures (McPherson & Hersch, 2000). In the present study, when the TLFB (IARA) was considered the reference standard, anonymous survey (CAHS) sensitivity was high for both any recent alcohol use and binge drinking, while specificity was unacceptably low for both any alcohol (0.54) and binge drinking (0.45).

A recent systematic review evaluating the accuracy of unhealthy alcohol screening tests in general primary care populations found that brief screening instruments (e.g., SASQ, AUDIT-C) reported sensitivity and specificity between 0.73 and 0.10; For the AUDIT-C, while sensitivity was similar, the range of reported specificity was wider. Most studies used structured diagnostic interviews as reference standards, with the TLFB sometimes being used in combination with the interviews. All of these studies were conducted largely in primary care settings and participants were administered a brief screener to identify excess use, followed by assessment with a more detailed instrument with greater specificity (e.g., the AUDIT) (O'Conner et al., 2018).

Interestingly, when O'Conner and colleagues (2018) looked separately at a sub-sample of adults with heavy use episodes (e.g., binge drinking) in the past month, PPVs ranged between 55-76% and NPVs ranged between 88-96%, results similar to the present study that found a PPV of 55% and NPV of 87%. In general, however, these studies evaluated the accuracy of screening tests in general primary care populations with a broad range of drinking disorders and patient characteristics, making it difficult to infer whether the test performs equally well in each group.

Similarly, what cannot be gleaned from present study findings, is whether the TLFB was, in fact, more valid than the CAHS in identifying alcohol and binge drinking or if alternatively, participants were more willing to admit to recent alcohol and binge drinking when queried through an anonymous, computer-delivered survey as compared to an in-person, face-to-face interview.

Notwithstanding the CAHS identified half as many non-binge drinking days (29%) as the IARA (57%), the anonymity, as well as method of administration, may have also resulted in higher self-disclosure of unhealthy use. Research suggests that heavy alcohol use is highly

stigmatized (Bazzi & Saitz, 2018) and that computerized surveys lead to significantly more reporting of socially undesirable behaviors than comparable surveys administered on paper (Gnambs & Kaspar, 2015) and during face-to-face interviews (Butler et al., 2009). In the CAHS modality, questions about unhealthy alcohol use were self-administered without an interviewer, allowing patients to respond privately and thus reducing the threat of social desirability bias (Davis et al., 2010; Estes et al., 2010; Richter & Johnson, 2001). As such, participants may have been tempted to give more importance to periods of lower alcohol consumption during the IARA. Further, order effect may have contributed to differences in response rates. Research suggests the orders in which are presented may affect how respondents answer subsequent questions (Strack, 1992). Again, for the present study, the anonymous, computerized survey was administered prior to the personal interview.

Additionally, offering certain option choices may affect test characteristics, albeit only to a small degree (Roy et al., 2009). For example, the present study asked, "During the past 30 days how many times (i.e., days) have you had 5 or 4 [males or females, respectively] more drinks per occasion?" while the TLFB recorded daily alcohol quantity and frequency without option choices. Interesting, previous research has validated a computerized single-question screener against a past-30 day TLFB for identifying unhealthy alcohol use among adults in primary care ("*How many times in the past year have you had X or more drinks in a day?",* where X is 5 for men and 4 for women) (McNeely et al., 2015b), demonstrating the validity of a clinically useful computerized brief screener.

Also, noteworthy, the categorization of binge episodes on the CAHS provided a less-than-perfect comparison to the TLFB. For example, days greater than 10 were categorized 11-20, 21-29, and every day (30 days) on the CAHS versus a continuous record of each day on the

TLFB. While the absolute difference is small (e.g., using TLFB data as the standard, the difference between methods in reports of binge drinking between 11-29 days was 8% on the IARA and 6% on the CAHS), the difference may be meaningful given the relatively large sample size of 540 participants.

Other methodological issues also deserve consideration. For example, was the definition of "a standard drink" adequately communicated during CAHS administration and did participants remember to consider beer and wine as alcoholic beverages (not just hard liquor)? Participants were offered a visual aid with standard drink definitions, but these definitions did not appear on the computer screen each time questions were asked. In contrast, TLFB interviewers routinely provided drink definitions (see Appendix A), and the TLFB was administered only to people RAs knew would be positive for alcohol and/or other drug use. RA's expected no completely negative cases, and interviewers could remind participants as they worked actively to assist with accurate recall (Sacks et al., 2003).

Alcohol Quantity and Frequency

For quantity-frequency of alcohol use per week (average drinks per week), a different pattern was observed, with participants reporting a greater number of drinks per week on the IARA (2.63) when compared to the CAHS (2.21). While past studies (Roy et al., 2008; Sobell & Sobell, 2003; Sobell et al., 2003) comparing quantity-frequency and daily estimation (DE) measures, such as the TLFB, have found relatively similar reports for aggregate drinking variables, quantity-frequency measures typically are not able to capture sporadic and atypical drinking patterns. Consequently, quantity-frequency measures compared with DE measures usually underestimate consumption. Further, in clinical trial data, the correlation between the

quantity-frequency and TLFB (using a 30-day recall period) is less robust (NIAAA, 2018), as most screening instruments assess for past year use (O'Connor et al., 2018).

In the present study, the agreement between quantity-frequency on the CAHS and the IARA (i.e., the TLFB) was moderate to good; however, more drinks per week were reported on the IARA vs. CAHS while more unhealthy/heavy alcohol use was reported on the CAHS when compared to the IARA. This combination of results has not been compared in previous studies.

The reasons for this discrepancy are not known and could relate to the different levels of data protection or due to the TLFB's sensitivity to sporadic and unpatterned drinking (Sobell & Sobell, 1992; Sobell & Sobell, 2003; Staudt et al., 2018). Further, determining quantity of alcohol use is complex. For example, respondents might think vaguely about recent drinking or they might carefully try to recall the number of drinks consumed. Subsequently, participants calculate their average number to format tentative answers in terms of the response options provided, or not provided in the case of the TLFB. Participants then decide what answer to choose based on social desirability bias. Collectively, this resulted in a much poorer crosswalk between quantity overall compared to the crosswalk between days of use and days of binge drinking in the past 30 days.

Interestingly, research reviewing the accuracy of quantity-frequency alcohol screening tests in primary care found a wide range of reported sensitivities and specificities (Fiellin et al., 2000). For example, one study found a sensitivity of 47% and a specificity of 96%, with the use of MAST scores as the reference standard, and a quantity cutoff score of 4 or more drinks per day (Cry & Wartman, 1988). Fleming and Barry (1991) found sensitivities of 50% and 20% with specificities of 87% and 97%, with the use of a cutoff of 7 and 20 drinks per week, respectively. In another study, there was a gradual decrease in sensitivity (100%-21%) with a corresponding

increase in specificity (43%-97%) as the number of drinks consumed per week increased from 0 to 24 or more (Buchsbaum et al., 1995). While this review took into account demographic and clinical (e.g., severity of alcohol problem) factors, these data suggest using formal screening instruments may perform better than quantity-frequency questions.

Of note, quantity-frequency of alcohol use per week was taken from the 30-day TLFB. The concordance between longer versus shorter assessment windows has been a subject of prior research (Carey et al., 2004; Hoeppner et al., 2010; Roy et al., 2008; Sobell et al., 2001; Toll et al., 2008; Vakili et al., 2008) but rarely investigated in non-substance-related treatment-seeking individuals. Research shows that participants report more drinking on the repeated TLFB-7 than on the standard TLFB-30, and those discrepancies between methods increased as the length of recall increased, suggesting that using shorter recall periods may yield more accurate data (Vakili et al., 2008).

Lastly, all consented participants in the present study were administered the TLFB assessment regardless of their CAHS results. This prevented workup bias (Reid et al., 1995), which occurs when participants preferentially receive the criterion standard evaluation based on positive, as opposed to negative, results on a screening test. When participants screen positive and preferentially receive the criterion standard evaluation, the sensitivity of the test can be falsely elevated because of the incorrect exclusion of participants (false negatives) from analyses. In addition, most, if not all, studies examining the accuracy of alcohol screening measures do so among general primary care samples and not among those with very heavy use. While the present study's sample included those with heavy/problem alcohol use, it also included those with heavy/problem illicit drug and/or prescription drug misuse. Because alcohol screening identifies patients with dependence (Saitz, 2010), it would be reasonable to argue that the

accuracy of the present study's screening tools would increase if the sample only included heavy/problem drinkers.

Other Drugs

For illicit drugs, when CAHS was compared to IARA, more participants also reported any recent drug use by personal interview (47% by IARA) than by anonymous, computerized survey (37% on CAHS). For prescription drug misuse, both the CAHS and IARA found 27% of participants reported one or more days of drug misuse in the past 30 days. The mean number of days of illicit drug use and prescription drug misuse in the past 30 were highest on the IARA. Participants also reported more average illicit drug use (1.44 by IARA and 1.42 by CAHS) and prescription drug misuse per week (0.68 by IARA and 0.39 by CAHS, $p < 0.001$) by personal interview; however, the difference between the CAHS and IARA for illicit drug use per week were nonsignificant.

Interestingly, reliability was low in detecting use of prescription drug misuse in the present study, however, even though the difference between the CAHS and IARA was statistically significant, with participants reporting the most days of prescription drug misuse on the IARA, in absolute terms the difference was minuscule. The relatively low reliability of screening for this substance class could be due to confusion among participants about what constitutes non-medical use (McNeely et al., 2014b). Also, there was no visual aid offered for illicit or prescription medications, and certain prescription drugs (e.g., sedatives and hypnotics) purchased on the street may have been mixed with illicit non-prescription drugs to create inconsistent recordings during the IARA.

When the IARA was considered the gold standard for assessment of illicit drug use,

CAHS sensitivity and specificity was 0.71 and 0.95, and CAHS's high PPV (0.91) is desirable, implying that false positive outcomes are minimized. For prescription drug misuse, CAHS sensitivity and specificity were slightly lower. In a recent systematic review assessing the accuracy of different screening tools (e.g., SUBS, TAPS, ASSIST, DAST) in adult primary care patients found the sensitivity of direct tools for detecting unhealthy use of any drug (including illicit drugs and prescription drug misuse) in the past month or year ranged from 0.71 to 0.94 (95% CI, 0.62-0.97), and specificity ranged from 0.87 to 0.97 (95% CI, 0.83-0.98). Screening tools had higher sensitivity for detecting unhealthy illicit drug use related to any drug (most of which was cannabis), cannabis, heroin, and stimulants than for detecting illicit drug use related to prescription drug misuse, including opioids or sedatives (range, 0.38-0.96 [95% CI, 0.29-0.99]) but specificity was comparable (range, 0.79-1.00 [95% CI, 0.71-1.00]) (Patnode et al., 2019; Patnode et al., 2020; USPSTF, 2020). One likely reason for the lower sensitivity in the present study is that many of the studies examining screening tools for other drug use assess use in the past three months. Research has shown that self-report rates of drug use increase when the recall window is widened to 90 days (Rendon et al., 2017).

While prevalence rates are similar to those found in other high-risk primary care populations (Saitz et al., 2014a), present study findings were unexpected as research has found computerized screening as well as provision of anonymity increases disclosure of substance use (Beck et al., 2014; Durant et al., 2002; Tourangeau & Yan, 1997). Further, research indicates that participants are more likely to disclose risky behaviors when assessed by Audio Computer-Assisted Self Interview (ACASI) format than by interviewer-administration (Dolezal et al., 2012; Estes et al., 2010; Perlis et al., 2004; Newman et al., 2002). A study examining audio ACASI versus facet-to-face interviewing found that the ACASI respondents were more likely to

report intranasal heroin use and smoking marijuana than face-to-face interviewing respondents (Perlis et al., 2004). Whereas this study examined interviewer-administered and computer-assisted modalities, their population included injecting drug users entering drug treatment programs, and the authors assessed for each illicit drug separately over the past six months. Nonetheless, this study highlights how greater privacy in ACASI conditions could have reduced social desirability bias, leading to greater and more accurate reporting of drug use (White et al., 2007).

The illegal status of drugs may underlie this distinction, where misuse of alcohol is more socially normative, especially in non-treatment settings. Relatedly, a major advantage about the interviewer-administered TLFB modality is that the face-to-face, interviewer-administered, semi-structured format establishes trust and rapport between interviewer and participant (Bowling, 2005; Rosenbaum et al., 2006), overcoming social desirability bias. Further, interviewers were not blinded to the study's inclusion criteria when they administered the TLFB, which could have impacted the interview. Even within surveys, differences in interviewer styles and presentation may influence validity.

Notably, willingness to disclose substance use can also be influenced by awareness of confirmatory biological testing (Nordeck et al., 2020). In the present study, participants agreed in the RCT consent to urine drug testing after completing the CAHS. This was done purposefully, to obtain responses that were not potentially biased by knowledge of subsequent testing, given the primary purposes of the parent study. Contrary to expectations, however, participants may have been more likely to provide accurate information after knowing their use could be discovered via confirmatory testing. This may also suggest that ensuring confidentiality is likely to maximize the truthfulness of self-reported drug use (Harrison & Hughes, 1997; Hjorthøj et al.,

2012; Magura & Kang, 1996; Shupp et al., 2020), as there are fewer consequences associated with illicit substance under research conditions. As a consequence, however, this finding might not apply to clinical assessment settings and may not generalize to less severe forms of use found among patients attending primary care clinics.

Nevertheless, these findings in general add to the existing literature supporting the use of the TLFB for dichotomous detection of substance use, and for frequency of such use (Hjorthøj et al., 2012; Martin-Willet et al., 2020; Metrik et al., 2018; Patnode et al., 2019; Patnode et al., 2020).

Anonymous vs. Confidential Survey (CAHS and CARA)

For alcohol, participants were more likely to report any recent (past 30-day) alcohol use by CAHS than by CARA. A similar pattern was observed for days of alcohol use in the past 30 days as well as average drinks consumed per week in the past 30 days. When the CARA was considered the gold standard, CAHS sensitivity was high for any recent alcohol use but specificity was unacceptably low, and reliability was moderate for both days of alcohol use in the past 30 and average drinks consumed use per week. One factor contributing to this finding may be variability in question wording as well as the window of assessment (week vs. month). However, CAHS had a high PPV (0.95) and moderate NPV (0.68), implying that false positive outcomes are minimized, and the false negative outcomes associated with moderate NPV is acceptable given the CAHS acted as a screening tool while the CARA offered a more detailed assessment of alcohol use.

Since actual alcohol use is unknown, we cannot determine if one computerized survey is more valid than the other. Given higher rates by anonymous vs. confidential surveys, it is possible participants were more willing to admit to recent alcohol use when no personal

identifying information had been collected. Similar to Beatty et al., (2014) who found participants disclosed significantly more alcohol use in an anonymous vs. traditional consent condition, Hormes and colleagues (2012) found an anonymous questionnaire yielded significantly ($p<.05$) higher rates of alcohol use compared to confidential survey in a sample of HIV positive primary care patients.

Whereas several computerized screening programs have been developed to identify alcohol use, research examining anonymous versus confidential computerized surveys is limited. Nevertheless, present study findings parallel those of previous studies in terms of the clear influence of anonymity on disclosure (Lau et al., 2003; Ong and Weiss, 2000).

Findings were mixed when comparing rates of other drug use. Interestingly, the highest rate of any recent (past 30-day) illicit drug use was reported on the CARA, and rates of prescription drug misuse were identical for both surveys. However, days of drug use and average days per week in the past 30 were higher on the CAHS compared to the CARA. In contrast, days of prescription drug misuse and average days of misuse per week in the past 30 were higher on the CARA (all ps $<.05$).

While these results were unexpected, they mirrored results reported by Beatty et al., (2014), who found that the presence of a certificate of confidentiality yielded increased disclosure of drug use compared to traditional confidentiality. Responding under confidential research conditions in the present study, with IRB consent language assuring protection against medical or legal consequences, may have provided some measure of comfort concerning the risks of drug use disclosure.

Further, consistent order effects and knowledge of biological testing could have influenced the truthfulness of their responses, particularly for other drug use. Finally, the CARA

provided a much more detailed assessment of alcohol and other drug use compared to the CAHS and also included a longer window of assessment (i.e., past three months). This provided a less sensitive comparison of past 30-day other drug use compared to the CAHS, and research suggests the validity of self-report measures of other drug varies when assessed for recent and more distal use (Cólon et al., 2002; Harrison & Hughes, 1997; Rendon et al., 2017), with validity increasing when the recall window is widened to 90 days. The influence of questions occurring earlier in the CAHS with subsequent, albeit differently worded, questions on the CARA could also account for the differences in self-reports (McNeely et al., 2014b). Interestingly, results obtained in the present study contradict conclusions drawn by researchers examining the role of confidential versus anonymous self-reports of substance use in other populations (e.g., younger adults; Moore & Ames, 2002; O'Malley et al., 2000), suggesting that findings may be unique to this patient group and setting.

Computerized Confidential Survey and Personal Interview (CARA and IARA)

Results were much more consistent when comparing the CARA and IARA, albeit findings contradicted hypotheses predictions. For alcohol, recent use and average drinks consumed per week in the past 30 days were higher on the IARA compared to the CARA; mean days of alcohol use in the past 30 were higher on the CARA, though, these differences were modest. For other drugs, mean days of drug use in the past 30 and frequency of drug use and prescription drug misuse were higher on the IARA compared to the CARA (all *p*s > 0.05). In contrast, any recent drug use was highest by CARA versus IARA and any recent prescription drug misuse was identical for both measures.

Interestingly, when the IARA was considered the gold standard for assessment of illicit drug use, the CARA had the highest sensitivity (0.89) and specificity (0.94). The higher

sensitivity and specificity for the CARA is likely related to the more detailed assessment of each illicit drug used in the past 3-months. Although it is desirable to have tests with high sensitivity and specificity, this result may not be of particular importance considering the length of the confidential, computerized assessment (Trevethan et al., 2013), which may not be useful in clinical practice settings.

While previous work has examined interviewer-administered and computer-assisted modalities in heterogeneous samples of patients, determining agreement between these two modalities has received less attention in the area of substance use in primary care settings. Moreover, while the TLFB has been validated in multiple formats, to date very few systematic comparisons have been conducted between computerized assessment and an interviewer-administered TLFB for substance use.

One study comparing interviewer-administered TLFB versus ACASI format questions about frequency of illicit drugs found the ACASI modality reflected more days used than TLFB (Delker et al., 2016). Specifically, they found a mean of 10.2 days by ACASI versus 8.7 days by TLFB. They also found the difference was greatest among younger participants. Similar to the present study, Delker and colleagues (2016) found that while statistically significant the magnitude of the discrepancy was marginal.

In contrast to the present study, Delker et al., (2016) assessed frequency of use during the prior 30 days by primary drug and included a sample of HIV-infected drug users. Substance use can have major consequences among HIV patients, which may have contributed to under-reporting on the interviewer-administered TLFB. Also, while the present study focused on the participants' primary drug on the TLFB, the CARA included all drugs the participant self-reported using, which may negatively skewed the distribution.

The present study expected that participants will report higher rates of alcohol and other drug use with a computerized self-administered assessment than interviewer-administered assessment (Beck et al., 2014; Butler et al., 2009; Delker et al., 2016; Lessler et al., 2000; McNeely et al., 2016a; Newman et al., 2002; van Griensven et al., 2006), however, several studies examining the effects of computerized self-administered surveys (CASI) have generated contradictory findings. Some comparisons of CASI with face-to-face interviewing have concluded that participants report more socially undesirable behavior in the face-to-face interview modes than with CASI (Tourangeau & Yan, 1997), while others have found little or no difference between CASI and face-to-face interviews (Islam et al., 2014; Williams et al., 2000). However, in particular, these studies include samples of injecting drug users from a Hepatitis B Acceptability and Vaccination Incentive Trial (HAVIT) (Islam et al., 2014) and drug users from an HIV reeducation risk study (Williams et al., 2000); participants may have been less motivated to conceal substance use from interviewers given incentives for study participation.

Similar results were found by McNeely and colleagues (2016a) who evaluated the concordance of an audio computer-assisted self-interview version of the Alcohol, Smoking, and Substance Involvement Screening Test (ACASI ASSIST) with the previously validated interviewer-administered ASSIST in primary care patients. The authors found the ACASI ASSIST demonstrated excellent concordance (92–99%) with the ASSIST in identifying moderate to high-risk substance use, though, reporting of alcohol and illicit drug use on the ACASI ASSIST. More reporting of illicit drug use CARA vs. IARA was also found in the present study, which is consistent with multiple prior studies showing that self-administered instruments generate higher rates of reporting of stigmatized behaviors (Kim et al., 2008). Similar to the present study, McNeely et al., (2016) summed ASSSIT responses for each

substance and also aggregated responses into two summary categories: ‘prescription drugs’ (prescription opioids, sedatives, and stimulants) and ‘illicit drugs’ (all other drugs, excluding tobacco and alcohol). However, their sample included a general medical population and relatively few had high-risk use of any substance. Further, participants completed both instruments in sequence, which has the potential to bias responses.

Possible explanations for the present study findings may include limited retrospective recall (Bowling, 2005; Killeen et al., 2004) on the CARA as it assessed for lifetime and past 3-month use. In contrast, the CARA preceded the TLFB and may have acted as a memory prompt, leading to increased reporting of substance use during subsequent TLFB administration. To eliminate this potential bias, future research should engage a cross-over design, in which half of the participants complete a CARA first while the other half begin with interviewer-administration.

Other factors contributing to these findings may be due to differences in measurement and terminology among the instruments (Bowling, 2005). For example, questions on the CARA focused on regular alcohol and other drug use in the past 3-months while the TLFB (IARA) focused on past 30-day quantity-frequency of use (Sobell, 2003), allowing only approximate comparability. Also, the TLFB allows interviewers to provide additional instructions and clarifications as needed. Further, participants responded under confidential research conditions in both formats, with the knowledge that the results would not enter their medical record and they were protected from legal of other consequences. One advantage of the TLFB is that the face-to-face, interviewer-administered, semi-structured format establishes trust and rapport between interviewer and participant (Rosenbaum et al., 2006).

Overall, differences between the present study and others are likely to reflect the characteristics of the sample, for example, whether the sample is in treatment or out of treatment, and the prevalence of use in the sample. Again, while it is unclear whether one modality is more valid than the other, results from the present study indicate that the interviewer-administered quantity-frequency measure produced more self-disclosure of alcohol and other drug use compared to the CARA. However, it is not known how patients' willingness to disclose substance use might change under real-world practice conditions.

Discrepancies Between Self-Reported Alcohol and Other Drug Use

Significant CAHS and IARA discrepancies in reports of recent alcohol and other drug use frequencies by some participants warrant further study. Some participants reported 30 days of other use on one measure but reported zero days of use on the other and vice versa. In general, these participants reported similar rates on the CARA or endorsed inconsistent use of other substances across measures, and many with discrepant prescription drug misuse data were recently in a controlled environment. There were also discrepancies between some participants reporting different frequencies of days of alcohol use and binge drinking in the past 30 days on the CAHS. Generally, these participants also underreported their frequency of binge drinking on the IARA, suggesting again the anonymity, as well as method of administration, may have resulted in higher self-disclosure of unhealthy use.

While some discordance between methods is expected related to methodological differences or random measurement error, social desirability bias likely contributed to the observed discordance. Participants may misreport substance use due to stigma or a desire to avoid further discussing their use (Davis et al., 2010). Furthermore, randomized controlled trials suggest that repeated screening leads to lower reported consumption on later screens

(McCambridge & Day, 2008; Moos, 2008; Kypri et al., 2007). The present study had a consistent order of administration, with provision of anonymity decreasing as participants progressed across the CAHS, CARA, and IARA. Further, participants could have experienced respondent fatigue (Lavrakas, 1998), especially if the individual participated in the RCT after a length medical appointment.

However, participants may also be more motivated to report substance use in subsequent surveys if they feel it is relevant to their health (Bradley et al., 2011). For example, participants became aware of the intervention/treatment focus of the parent study during confidential consent procedures and may have been more honest with their use during the IARA if interested in addressing their use. For example, previous research by members of our research team found those consenting to the RCT endorsed a range of mental health conditions and greater sleep disturbance than those who chose not to participate (Kelpin et al., 2018). This finding may reflect the association between help-seeking and medical comorbidity (Aikens & Rouse, 2005; Hall & Farrell, 1997). Greater awareness of the need for help among clients with comorbidities might make them more motivated to seek services, including those provided through a clinical trial.

Also, the primary purpose of the CAHS was to identify those at risk for heavy/problem substance use, this was not made explicit during recruitment. Instead, it was described as an anonymous general health and risk behaviors survey. This was done in part, to minimize social desirability. Therefore, the survey included questions that did not pertain to substance use. Research efforts were made to mask the primary purpose of the screener, however, may have become less successful over time, as the clinic became aware of an ongoing RCT that involved substance use. Patients could converse in the waiting area and this may have influenced how they responded to the substance-related survey items.

While the number of cases with discrepant data were small compared to the larger sample (e.g., 3-7 cases per comparison), these data highlights that measuring frequency of alcohol and other drug use can be vulnerable to under-reporting or misreporting likely due to different levels of data protection and social desirability bias as well as other factors that are unpredictable based on this study.

Identifying Alcohol and/or Other Drug Use Related Problems

The present study examined the concurrent validity of the ASSIST-3 in comparison to the CAGE, T-ACE, and CAGE-DRUG. The optimal cut-points on the ASSIST-3 for detecting alcohol-related problems on the CAGE/T-ACE and for detecting illicit drug use-related problems on the CAGE-DRUG were 1 and 2, respectively. The ASSIST-3 was found to have acceptable sensitivity and specificity to detect alcohol-related problems in both males and females. The ASSIST-3 had good sensitivity but poor specificity for detecting illicit drug use-related problems in males, and the ASSIST-3 had acceptable sensitivity but poor specificity for detecting drug use-related problems in females.

The present study also compared positive scores for problematic substance use on the CAHS (i.e., CAGE, T-ACE, and CAGE-DRUG) by CARA (i.e., ASSIST-3). In general, both males and females were more likely to score positive for alcohol-related problems on the CAHS compared to the CARA. These results were in the predicted direction. While females were more likely to score positive for illicit drug use-related problems on the CAHS, males were more likely to score positive on the CARA.

The existent literature supports the usefulness of screening for both alcohol and drug use disorders and their antecedents in primary care (Pilowsky & Wu, 2012; USPSTF, 2018; USPSTF, 2020), and the present study offers insight on the performance of a modified ASSIST

to screen for substance use-related problems. While sensitivity, specificity, PPVs, and NPVs were mixed, the ASSIST-3 could prove useful by identifying individuals with current alcohol and/or drug use and problems. In contrast to the CAGE, T-ACE, and CAGE-DRUG, the ASSIST-3 inquiries about specific substances and assesses for current problems. It is much briefer and easier to score than the full ASSIST and has the potential to lead to actionable results based on the scores, such that individuals who score positive need an assessment for treatment. However, additional refinement and validation against alternative criterion measures (e.g., the full ASSIST; AUDIT, CIDI; TAPS Tool; DAST; DUDIT; Inventory of Drug Use Consequences (InDUC) (Tonigan & Miller, 2002)) will be important before it can be broadly recommended as a screener for substance use-related problems.

For example, the ASSIST-3, in contrast to the CAGE, T-ACE, and CAGE-DRUG and in keeping with the format of the ASSIST, has three items for each substance and an ordinal response format rather than a binary (yes/no). Based on previous literature the ASSIST-3 questions were consolidated (McPherson & Hersch, 2000; Tiet et al., 2008; Tiet et al., 2015) and each category of drug (e.g., marijuana, cocaine, heroin) was assessed as a combined question, with the final score (i.e., for illicit drugs) based on the highest value for each question. This format as opposed to dichotomous (yes or no) responses provides information about the specific substances used and severity of problems to guide clinical actions. However, this format likely limited the sensitivity and specificity of the ASSIST-3.

Similar to Steinbauer and colleagues (1998), normal cut-points were used for the CAGE (2), T-ACE (2), and CAGE-DRUG (1) questionnaires. Using different cut-points for the reference standards could have also impacted results. For example, while the Consensus Panel recommends that the primary care clinicians lower the threshold to one positive answer to cast a

wider net and identify more patients who may have substance abuse disorders (Center for Substance Abuse Treatment, 1997), the present study included a unique sample of individuals who met criteria for heavy/problem substance use. Lowering, or rising the threshold cut-points may have led to an underestimation of CAGE sensitivity but would have likely excluded a large number of individuals and limit the ability to draw some comparisons between the CAGE, T-ACE, CAGE-DRUG and ASSIST-3.

This is first study to examine the concurrent validity of a modified ASSIST focused exclusively on problems. While the ASSIST-3 was likely a better assessment of current substance use-related problems compared to the CAGE, T-ACE, CAGE-DRUG, the present study did not provide additional criterion measures to compare. Further, there is limited ability to draw comparisons with other instruments that screen for substance use-related problems, because only the ASSIST, ASSIST-Lite, and the TAPS Tool (a modified version of the ASSIST-Lite) provide substance-specific results.

In comparison to the TAPS Tool (Schwartz et al., 2017), the ASSIST-3 underperformed. The TAPS Tools, compared to the WHO ASSIST, had favorable sensitivity and specificity at a cut-point of 2 to detect high risk use of tobacco, alcohol, marijuana, stimulants (prescription and cocaine/methamphetamine combined) and opioids (prescription opioids and heroin combined), but its sensitivity was unacceptably low in detecting moderate risk use of stimulants (cocaine, prescription stimulants) opioids (heroin, prescription opioids), and sedatives at a cut-point of 1. While Schwartz and colleagues (2017) conducted their research anonymously, participants were included all primary care patients while the present study looked at a unique higher risk subgroup that enrolled in a RCT. Also, the TAPS Tool's items mapped onto the WHO ASSIST

classifications. These factors likely contributed to the differences in sensitivity and specificity between the ASSIST-3 and TAPS Tool.

It is possible that the accuracy of the ASSIST-3 also reflects differences between the timeframe of the reference standard measure and the ASSIST-3. The ASSIST-3 screens for problems in the past 3 months (vs. lifetime with the CAGE, T-ACE, and CAGE-DRUG). As a result, the ASSIST-3 could fail to identify individuals who had problem use in the past (e.g., concern expressed by friends or relatives, and failed attempts to control, cut down, or stop using) that had not continued into the most recent 3-month period. However, by focusing on current use, the ASSIST-3 has the potential to identify individuals who are most in need of clinical intervention, which is important in primary care settings. In contrast, participants who did not use the substance in the past 3 months would have a zero score on the ASSIST-3.

This is also the first study to compare an anonymous versus confidential self-administered computerized screening tool. Endorsement of "problem use" for alcohol and other drug use was higher on the CAHS compared to the CARA. One likely reason for this is finding is that participants are more reluctant to disclose substance use-related problems in a confidential vs. anonymous survey. However, and like mentioned before, another likely reason for this finding is that participants who reported lifetime substance use denied used in the past 3 months would not be identified by the ASSIST-3.

Interestingly, and in contrast with previous research (Kypri et al., 2005), a greater proportion of females screened positive for at-risk alcohol use relative to males. The ASSIST-3 also performed best when compared to the T-ACE for alcohol (AUC = 0.76). One likely reason for this finding is that the present study had a larger sample of females compared to males. The present study also used a modified version of the T-ACE (by assigning only 1 point to the

Tolerance question). Present study results concerning alcohol-related problems in women are similar to a cross-sectional study examining prevalence of symptoms alcohol abuse (McQuade et al., 2000) in a hospital-based outpatient clinic. The authors, who also modified the T-ACE by assigning only 1 point to the Tolerance question, found that the T-ACE was more effective than either the CAGE or the AUDIT in identifying patients who meet diagnostic criteria (sensitivity) and distinguishing patients who do not meet criteria (specificity). One important limitation of this finding, however, is that these results cannot compare directly to clinic data.

While males were more like to score positive on the ASSIST-3 vs. the CAGE-DRUG, the ASSIST-3 outperformed the CAGE-DRUG in males compared to females. Research shows that dependence on or harmful use of illicit drugs is generally higher in men (National Institute on Drug Abuse [NIDA], 2020) compared to women, though, women are at greater risk for a variety of medical and psychosocial consequences (Polak et al., 2015). However, research is limited and mixed on gender differences in screening for drug use and related problems in primary care. The single-item screen (How many times in the past year have you used an illegal drug or used a prescription medication for non-medical reasons?) developed by Smith and colleagues (2010) was found to be less specific for the detection of a current drug use disorder in men compared to females, but the difference was small. The Substance Use Brief Screen (SUBS), which screens for all clinically relevant classes of substances (tobacco, alcohol, illicit drugs, prescription drugs used nonmedically), had statistically significant lower sensitivity and higher specificity among females for detecting unhealthy use of any drug (McNeely et al., 2015a). Future studies should examine how well screening tools perform by gender, especially tools screening for illicit drug-use related problems.

The USPSTF currently recommends screening for drug use in primary care, and the USPSTF review found no evidence addressing harm from drug screening. Continuing to screen in primary care is important as it reduces stigma and invites substance use to be viewed as an important and treatable health problem. Notwithstanding, any screening tool assessing for risk associated with unhealthy alcohol or drug use or comorbid conditions only reveal information signaling the need for further assessment. Therefore, brief screening tools that inquire about recent use of specific substances and goes beyond simple endorsement of use (e.g., TAPS Tool, ASSIST-LITE, AUDIT-C) can lead to more actionable results versus tools like the CAGE, T-ACE, CAGE-DRUG, or single substance screeners such as the AUDIT for alcohol (Bradley et al., 2003) that provide less data on problematic use.

Agreement Between Self-Reported Recent (past 30-day) Other Drug Use and Urinalysis

Overall prevalence of positive UDS results and endorsement of use on the TLFB included: marijuana (30% positive UDS, 33% positive TLFB), cocaine (20% positive UDS, 14% positive TLFB), heroin / opiates (20% positive UDS, 24% positive TLFB), opiates (20% positive UDS, 20% positive TLFB), benzodiazepines (16% positive UDS, 4% positive TLFB), heroin (13% positive UDS, 4% positive TLFB and amphetamines (2% positive UDS, 0.5% positive TLFB). The overall concordance between self-reported use on the TLFB and positive UDS results ranged from 72% for non-medical opiate use to 97% for non-medical amphetamine use; however, non-medical amphetamine use was not highly prevalent and the congruence of self-report and urine test data can be inflated if drug use is not highly prevalent in a sample (Harrison, 1997).

These findings are fairly consistent to those of a systematic review and meta-analysis examining the validity of the TLFB for illicit drugs in comparison to biological measures

(Hjorthøj et al., 2012). The included studies almost exclusively used urine as a biological measure and involved subjects with a diagnosis of alcohol or illicit substance use disorder. The authors found that the percent agreement between self-report and biological measures ranged from 87.3%-90.9% for marijuana, 79.3%-84.1% for cocaine, and 94.0 for opiates, evidencing the validity of the TLFB in detecting illicit substances in populations with substance use disorders. Further, Hjorthøj and colleagues (2012) found sensitivity and specificity were 0.60 and 0.42 for marijuana, 0.69 and 0.9 for cocaine, 0.88 and 0.96 for amphetamines, and 0.64 and 1.00 for opiates. In the present study, sensitivity (0.80) and specificity (0.87) were higher for marijuana, and while the sensitivity was much lower for cocaine and opiates in the present study, specificity and NPV were high for each drug, highlighting the TLFB's ability to validly detect true negatives.

The different sensitivity, specificity, and rates of agreement between the present study and that of Hjorthøj and colleagues (2012) are likely due to length of recall. In the review by Hjorthøj et al., (2012), the length of recall (i.e., the number of days of recall assessed by the TLFB) varied in many studies from the actual number of days used for validation with UDS samples. For instance, if the potential detection window for a substance in UDS was only five days, some studies only used the last five days of TLFB for comparison. However, Hjorthøj and colleagues (2012) as well as Rendon et al., (2017) found the agreement rate increased as the length of recall increased, suggesting that utilizing the full TLFB increased validity. As such, in the present study, the length of recall was the past 30-days regardless of detection window for a substance in the UDS.

Notwithstanding, this may have contributed to the low concordant opiate, heroin / opiate, and heroin self-report and urinalysis given the relatively short detection window of opiates (e.g.,

1-4 days; ASAM, 2017). Present study results are similar to that of Chermack and colleagues (2000), who compared self-reports using the Addiction Severity Index (a semi-structured interview assessing frequency of drug use (opiates, cocaine) in the past 30 days) (McLellan et al., 1992) and urinalysis. The authors found conditional k's (cond. k) for opiates were closer to chance agreement (cond. k's ranging from 0.07 to 0.24), suggesting that urinalysis tests were likely unable to identify drug use for a large proportion of the 30-day time interval assessed by the ASI. However, Chermack et al., (2000) compared self-reports to urinalysis among patients enrolled in methadone treatment, and comparisons across study should be made with caution regarding the degree of under-reporting under naturalistic conditions versus research conditions.

Furthermore, while the present study found overall high concordance across substance, nearly all of the concordance in this study was attributable to participants who indicated non-use on the TLFB and whose UDS results were negative. The low value of the (cond. k's) indicates that concordance positive rates were poor, with the exception of marijuana and cocaine (cond. k > 0.50) and suggests that results were likely influenced by the combination of patient underreporting and infrequent urinalysis tests.

Differences in results could also be related to different study populations. In particular, nonmedical use of prescription opioids and heroin is a growing problem in the U.S. (Compton, Jones, & Baldwin, 2016), and the stigma associated with opioid use may be more apparent in healthcare settings (Olsen & Sharfstein, 2014; Yang et al., 2017). Also, Black individuals in the U.S. have been subject to discrimination, and addiction to opioids is seen as a condition largely affecting these disadvantaged minorities (Carliner et al., 2016; SAMHSA, 2020b). The present study's sample consisted largely of Blacks (78%), and the intersectionality of race and stigma

could have contributed to the low concordant opioid self-report and urinalysis (White et al., 2014).

Overall concordance rates in the present study were somewhat lower compared to a recent study by Nordeck and colleagues (2020), who examined concordance of self-reported substance use using a past 30-day TLFB but compared to oral fluid testing (OFT). The study was conducted under research-confidential conditions in a sample of primary care patients. Specifically, the authors found overall concordance rates between TLFB and OFT were 94.9% or higher for each substance. Other research has also identified lower levels of agreement between biological testing and self-report (e.g., Preston et al., 1997; Rendon et al., 2017), and in his review of studies investigating treatment and non-treatment samples of drug users, Darke (1998) reported that concordance between urine testing and self-report data varied from 71% to 95%, similar to the present study.

Unlike the present study, Nordeck and colleagues (2020) conducted the TLFB prior to informing participants of the opportunity to provide an OFT, with the aim to prevent biased responses, and this may have affected the validity findings (Wish et al., 1997). For example, Hamid et al. (1999) demonstrated that agreement between urine test results and self-report of opiate and cocaine use increased from 58% to 93% when the urine tests were performed before the self-report interview. Nordeck and colleagues (2000) also limited TLFB data to the time period that aligned with the OFT detection window, which ranged from 1 to 3 days depending on the substance.

Interesting, Nordeck and colleagues (2020) found that marijuana had lower detection on OFT than self-report (27.6 % OFT-positive only vs 32.2 % TLFB-positive only), similar to the present study (30% UDS-positive only vs. 33 % TLFB-positive only). While it is unclear why

participants who self-reported recent marijuana use had a negative UDS, one possible explanation could be related to false negative UDS results. Research has also demonstrated that for marijuana, adherence (i.e., THC compounds migrating to non-aqueous surfaces such as the sides of bottles) may reduce the concentration of the drug in the urine specimen, perhaps below positive cut-off levels (Riley, Lu, & Taylor, 2000). It is possible that adherence may have occurred. Further, participants may be more willing to disclose marijuana use given its increasing prevalence and acceptance among the general public (Hasin, 2018); as a result, their use could have been inadvertently misreported.

In the present study, rates of heroin / opiate use were also higher by self-report (24%) compared to UDS (20%), though, the difference was modest. Possible explanations for this discrepancy could relate to length of recall or cross-reactivity with other substances as well as other factors associated with the UDS detection window (e.g., dosing, metabolism, body mass, urine pH, duration of use, drug pharmacokinetics, etc.) (Riley et al., 2000).

Present study findings were also fairly consistent with those of Neale and Robertson (2003), who reported a concordance rate of 85% across multiple substances (e.g., opiates, benzodiazepines, methadone and cannabis), and to Cone (2012) who reported a concordance rate of 95% for cocaine, but only slightly more than 50% for heroin use. However, in studies by Neale and Robertson (2003) and Cone (2012), data were collected via standardized questionnaires relating to drug use in the past 3 (Neale & Robertson, 2003) and 7 (Cone, 2012) days preceding baseline interview, self-report results were compared to OFT, and samples included participants from drug treatment facilities. While the present study included a non-treatment sample, study eligibility required endorsement of heavy/problem alcohol and/or drug use.

In general, findings suggest that relying solely on self-report would have missed a number of participants with evidence of cocaine, heroin, opiate, benzodiazepine, and amphetamine use as indicated by positive UDS results. Conversely, relying solely on UDS results would also miss evidence of substance use via self-disclosure (e.g., marijuana; heroin / opiates), which might vary markedly depending on the substance (Fendrich et al., 2003). Of note, the present study only included those who reported prescription drug misuse by TLFB. UDS testing does not discriminate between reasons for prescription drug use, and thus participants who did not report prescription drug misuse may be positive for prescription medications. This is one likely reason for the stark difference between TLFB-positive only (4%) and UDS-positive only benzodiazepine (16%) results.

Although it is not known what the cause of inconsistencies between patients self-reported illicit drug use and UDS was in this study, there are several possible explanations. Self-report data of recent illicit drug use can be flawed for numerous reasons including faulty memory (Darke, 1998), imprecise knowledge of the amount and purity of the drug, and fear of the consequences of admission of drug use despite assurances of confidentiality. Moreover, some substances are detectable in biological samples for periods that may exceed the detection window for TLFB, potentially leading to the classification of true-negative TLFB reports as false negatives. Conversely, some substances are detectable in biological samples only briefly, with true-positive TLFB reports being classified as false positives.

Since these findings compare with the results of similar studies based on urine testing, it seems reasonable to argue that TLFB is both a valuable tool for detecting recent drug use and a useful independent indicator of the validity and reliability of drug users' self-report data. That said, substance use behavior should ideally be measured by a combination of self-report and

biological indicators (Donovan et al., 2012). Also, future research should focus on replicating these results in primary care populations as well as further examining the variations in levels of concordance and discordance between the drug categories.

Exploratory Analyses

There were several differences observed in psychosocial correlates of self-reported substance use across each assessment. While there was no specific pattern, the most common psychosocial correlate for recent (past 30-day) alcohol use on the CAHS was being from a minority group; for the CARA it was being female and having government insurance, and for the IARA, it was having >12 years of education. The most common psychosocial correlates for recent illicit drug use on the CAHS were being older and endorsing moderate-severe psychological distress; on the CARA, they were trouble staying asleep in the past 30 days, being older and female, and endorsing moderate-severe psychological distress. On the IARA, only being older was the most commonly associated correlate with any recent illicit drug use. Taking sleep aids in the past 30 days and having 2+ medical conditions were the most common correlates associated with prescription drug misuse on the CAHS; on the CARA, trouble falling asleep in the past 30 days, being older age, and having 2+ medical conditions were most commonly associated with recent prescription drug misuse, while endorsing symptoms of anxiety was the most common correlate associated with prescription drug misuse on the IARA.

Much of the existing literature focuses on correlates of SUDs in primary care populations. While the present sample included heavy alcohol and other drug users, diagnosing with a formal instrument was not completed. As a result, the generalizability of present study findings to other primary care populations must be made with caution. Further, methodological differences (setting, interview mode, contextual effect) complicate the comparison of our results

with those from other survey or clinical trial samples. However, discussing present study results in the context of the existing literature provides opportunities to inform research in primary care patients at-risk for substance use disorders.

Present study findings are somewhat inconsistent with previous research (John et al., 2018; Wu et al., 2017). In a primary care sample of past-year substance users, John and colleagues (2017) found that individuals who were male, white, less educated, disabled, or not married had increased odds of having a single SUD, while participants who were male, ages 26–34, less educated, and unemployed had increased odds of multiple SUDs compared to one SUD. In a similar sample of 2000 adults, younger age (18-25), male sex, and low education were associated with increased odds of having a SUD (Wu et al., 2017). These results were despite a relatively large distribution of females, older adults, and blacks, similar to the present study. However, samples for these studies included primary care patients regardless of substance use while the present study included only those with heavy/problem alcohol and/or other drug use. The psychosocial characteristic comparison between the present study and the latter findings is important, however, since these results provide initial evidence that psychosocial differences exist between general primary care populations and heavy/problem substance users in primary care, which could impact other clinical outcomes (Rohn et al., 2017).

Other research has found that SUDs are associated with major medical conditions. For example, in a study of participants from a large integrative health care system, patients with SUDs had higher prevalence of 19 major health problems, with chronic pain, chronic obstructive pulmonary disease, congestive heart failure, and hepatitis C being among the most elevated (Bahorik et al., 2017). In another study of patients with co-occurring diabetes and hypertension, 1.9% of patients also had opioid use disorder, 2.2% had cocaine use disorder, 1.1% had cannabis

use disorder, and 8.8% had alcohol use disorder (Winhusen et al., 2019), and in a sample of adult patients with high risk diabetes, 48.3% had a comorbid SUD (Wu et al., 2018). Although these studies used data that were based on electronic health record (EHR) documented diagnoses that were provider identified, rather than comprehensively screened for or assessed. EHR data can be influenced by biases (e.g., misclassification, severity of medical conditions, provider specialty), and the recency of EHR-based diagnoses cannot be precisely defined (Wu et al., 2013).

Largely consistent with present study findings, previous research by members of our research team found present study-RCT consenters reported experiencing a larger number of psychosocial and medical comorbidities than non-consenters. In particular, older age, being unemployed or on disability, endorsing problems with anxiety and depression, and trouble falling asleep (past 30 days) were all associated with consenting to participate in the SBIRT RCT. Participants with ≤ high school degree, as well as those who were employed or retired was associated with declining study participation (Kelpin et al., 2018). In general, these patterns could be a result of SUD-related health problems requiring treatment (Pilowsky & Wu, 2012), and that participants were more ready to identify they are in need of services due to their comorbidities. Interestingly, recent research has also found that nontreatment seekers for alcohol use (compared to treatment-seekers), were more ethnically diverse, less educated, single, and working part-time or unemployed (p's < 0.05) (Haass-Koffler et al., 2020).

Comorbid substance use and mental health disorders have been shown to be associated with greater substance use severity (Gorfinkel et al., 2020; Grant et al., 2016). Interestingly, screening for substance use can also lead to the potential identification of these comorbid disorders. For example, Khan and colleagues (2020) found that within a sample of Veterans Health Administration patients, an AUDIT score of 20 or higher (vs <8, the reference) was

associated with symptoms of depression and anxiety. This highlights how alcohol screening alone has potential to convey substantial information regarding the likely presence of alcohol-clustering conditions, particularly for depression and anxiety, and inform decisions about additional assessment. Taken together, these findings reinforce the importance of promoting the integration of evidence-based mental health and substance use screening in routine medical settings, which currently are not used consistently (Bazzi & Satz, 2018; Edelman, & Tetrault, 2019).

It is also important to note that patients with SUDs may face stigma or other barriers to engaging with primary care (van Boekel et al., 2013), which can contribute to disparities in preventive care (Hirsh et al., 2020; Hoggatt et al., 2019). Despite considerable study on disparities in health care experiences by race/ethnicity and gender (Jones et al., 2016), continuing to explore other correlates of substance use and identify disparities is important for ensuring health care equity for all vulnerable groups. Furthermore, research on the driving forces contributing to these disparities is needed.

Study Implications and Applications

This study has a number of important implications. First, it provides benchmark data on comparing computerized anonymous and confidential surveys to a confidential, interviewer-administered assessment across alcohol and illicit drug use as well as prescription drug misuse. Overall, participants self-reported higher rates of substance use on the computerized, anonymous health survey (CAHS), including any recent (past 30-day) alcohol use, binge drinking, and problems associated with alcohol use. While rates of any recent (past 30-day) illicit drug use were higher on the CARA, this tool provided more in-depth assessment of use across several classes of illicit and prescription drugs and included a window of past 3-months. For quantity-

frequency of alcohol use as well as frequency of illicit drug and prescription drug misuse, self-reported use was highest on the IARA (i.e., the TLFB). While the CAHS had good sensitivity, specificity, acceptable agreement and reliability for identifying alcohol and illicit drug use and associated problems, for substances that are less frequently encountered in primary care, sensitivity and specificity estimates were lower and less precise. Nevertheless, research suggests that increased reporting of substance use is a sign of improved validity in the methodology since these behaviors are typically underreported (Beck et al., 2014).

Thus, while screening for recent alcohol and other drug use by the CAHS provides the most parsimonious method to screen for substance use and related behaviors in primary care, collecting the most accurate data on the substance(s) can be improved with additional follow-up using a confidential TLFB. This approach will help provide sufficient information to inform clinical practice and looks separately at difference substances (Gryczynski et al., 2017; Smith et al., 20010; Saitz et al., 2014b; Tiet et al., 2015), and follows the current recommended guidelines which suggest starting with brief screening (1-2 questions) with follow-up on cases where substance use was detected to collect diagnostic data (Sayre et al., 2020). It's important to note that the CAHS utilized CIAS and was avatar guided. This practice is becoming more relevant within the larger alcohol-SBIRT literature where several trials embed alcohol questions in a broader health and lifestyle assessment survey to reduce social stigma of substance use and tendency to under-report use (Blow et al., 2006; Cucciare et al., 2013; Dimeff et al., 2000; Haskins et al., 2017; Montag et al., 2015; Neumann, et al., 2006; Ondersma et al., 2016; Tzilos et al., 2011). Anonymous surveys should take this into consideration for future research and clinical applications.

Studies have shown that allowing people to answer questionnaires completely anonymously yields more reports of socially inappropriate attitudes, beliefs, and behaviors, however, anonymous research designs do not allow for longitudinal follow-up with linking of data across observations. Another challenge implementing anonymous surveys is connecting patient data to their medical record. Therefore, it may be impractical to try and ensure absolute anonymity, especially when the goal is translation to clinical practice. As such, one way to utilize an anonymous survey limit any privacy risk of data is by ensuring confidentiality and controlling access to the data (Beatty et al., 2014; Wayal et al., 2018).

This study also further documents and strengthens the evidence supporting the construct validity of the TLFB. While utilizing a CAHS followed by an interviewer-administered TLFB may be a useful way to capture sensitive and accurate data about substance use in primary care from a research perspective (Bobak et al., 2004; Rehm et al., 2001; Russell et al., 2004), it may not be the best clinical practice. For example, length of administration and other limitations (e.g., availability of clinic and research team resources) are major disadvantages of the TLFB. However, if the goal of a research study lies in comparing two groups in their volume of alcohol consumption, such level of detail may not be accurately captured by a CAHS. Therefore, continuing to utilize the TLFB to capture accurate data related to quantity and frequency remains best research practice.

Additionally, present study findings are similar with others comparing self-report with urine testing. While the TLFB can be a useful independent indicator of the validity and reliability of drug users' self-report data, the present study found that the TLFB and UDS showed disparate detection of different substances. These findings suggest that using a combination of self-report and biological indicators may improve detection of drug use (Donovan et al., 2012).

The economic implications of present study findings warrant consideration. Research shows that the cost of substance abuse treatment is substantial (Florence et al., 2018), and the presence of a substance use disorder often doubles the odds that a person will develop another chronic and costly medical illness (Scott et al., 2016). Further, recent examination of healthcare spending shows that individuals who have comorbid mental health and substance use disorders account for almost half of the annual total healthcare costs (Davenport et al., 2020). Primary care settings have been at the forefront of delivering substance use disorder services in mainstream health care, resulting in significant health-care savings (Babor et al., 2007; HHS, 2016). While primary care providers may utilize brief screeners for alcohol and other drug use, interviewer-administered screening approaches can be challenging to implement because they require staff time and training, and the cost of using a skilled interviewer for the TLFB is considerable (Maisto et al., 2008). Moreover, implementation cost estimates for screening in primary care indicate that other key cost drivers are service support costs for screening (Zarkin et al., 2015).

This has prompted the development of alternative administration methods, in particular, computerized methods (Delker et al., 2016; McNeely et al., 2016a; McNeely et al., 2016b). Embedding standardized, validated clinical assessments of substance use, including the TLFB (Martin-Willett et al., 2020), into self-administered electronic platforms may not only facilitate the assessment of substance use as part of the routine clinical workflow and may also be cost-effective (Gryczynski et al., 2017; Tai et al., 2012). Further, because the TLFB produces a valid description of patterns and amounts of illicit substance use, it holds practical cost advantages over the use of biological measures such as samples of urine, blood, or hair.

Lastly, several studies have focused on computerized modes of administration to counteract substance use-related stigma and address confidentiality concerns, demonstrating that

computerized self-administered instruments generate higher rates of self-reported substance use (Beck et al., 2014; Lessler et al., 2000; McNeely et al., 2016a; NIMH, 2008; Perlis et al., 2004; van Griensven et al., 2006), consistent with the present study. While attitudes within the clinical treatment system of individuals with substance use disorders are shifting, sociocultural factors associated with problematic drug-using populations, such as fear, lack of information and education, general physical and mental health problems, homelessness, and incarceration further stigmatizes people who use drugs, making it more difficult to engage people in health care and other services (American Public Health Association [APHA], 2013). Consequently, focusing on computerized screening and assessment will likely decrease stigma associated with alcohol and other drug use, enable widespread reach and scalability of evidence-based practices, and provide a more equitable distribution of health resources and existing racial and class-based inequities (Marsch et al., 2020).

Addressing the Dual Challenges of Substance Use Disorders and COVID-19.

The U.S. is currently facing an opioid epidemic (SAMHSA, 2019), and the Coronavirus disease 2019 (COVID-19) pandemic is projected to be the largest mass casualty event in U.S. history. The COVID-19 pandemic is having a wide range of negative impacts on people affected by a variety of health conditions, namely those with mental health and substance use disorders. Recent analysis of nationwide surveillance data found that drug overdoses rose by 18% in March, 29% in April, and 42 percent in May compared to the same months in 2019 (ODMAP; Alter & Yeager, 2020). This increase in opioid overdoses has been observed locally as well, where the total number of opioid overdose visits at VCUHS increased from 102 to 227 in the first four months of the COVID-19 pandemic compared to the same time period in 2019 (Ochalek et al., 2020). Additionally, a recent study examining alcohol consumption in the U.S.

during COVID-19 found an increase in adult consumption by 14% between spring of 2019 and spring of 2020. Also, heavy drinking episodes by women increased by 41% (Pollard et al., 2020).

Further, symptoms of anxiety and depression increased considerably during April-June of 2020, compared with the same period in 2019 (Czeisler et al., 2020). Again, this reinforces the importance of promoting evidence-based screening in routine medical care. Fortunately, the Centers for Medicare and Medicaid Serves have issued sweeping changes to make telehealth services easier to access as the COVID-19 crisis has made routine care more difficult to deliver. With the increase in health and telehealth communication strategies (Pierce et al., 2020), incorporating clinically meaningful screening tools into telehealth systems will likely help address these conditions associated with the COVID-19 pandemic.

Study Strengths, Limitations, and Future Directions

Strengths. The present study had a number of strengths. First, whenever possible, the CAHS used reliable and valid measures to assess for problems in each domain of interest. To avoid practitioner bias, every patient was asked the same set of questions (Svikis & Reid-Quiñones, 2003). Further, the format and delivery of the CAHS survey promoted patient anonymity and included many items irrelevant to substance use in an effort to mask the true purpose of the screener as well as to reduce the stigma associated with substance use. Also, the use of the self-administered electronic screener provided a comprehensive, reliable, private, and less biased approach to collecting data. Moreover, the Computerized Intervention Authoring System (CIAS) format was engaging and less intimidating than traditional health screeners. Lastly, the CAHS and CARA included branching logic that streamlined the screening tool and all answers were recorded by simply tapping responses from a list. This helped cut down on the overall time to complete the screener and limited missing data.

Second, the study included random assignment and produced approximately equal group assignments.

Third, the different level of data protection as well as method of administration across the three substance abuse assessment measures allowed for a novel comparison within a primary care population. The present study provides benchmark data on frequency of other drug use, adding to the literature suggesting the TLFB is a useful tool for assessing other drug use frequency patterns.

Fourth, the study included biological measures of substance use (e.g., urine drug screen), offering confirmatory measures of self-report data. Further, the study emphasized that the research study was independent of the patient's treatment and would not be shared with VCUHS staff, supporting participant confidentiality and overall comfort with study participation.

Limitations: Despite these strengths, the study also had a number of limitations. First, sampling was only from on primary care clinic in Eastern U.S. and included participants who only spoke English language. The clinic also serves predominately low-income, ethnic minorities. Hence, findings may not generalize to other parts of the U.S., to other countries, or be representative of patients seen in all primary care settings. Additionally, while RCT eligibility had few exclusion criteria, the present study sample only included those who met criteria for heavy/problem alcohol and/or drug use. Moreover, of those who met RCT criteria only half consented, and differences were found between those who did and did not consent to the RCT (Kelpin et al., 2018). For example, consenters were more likely than nonconsenters to report recent and more frequent drug use, prescription drug misuse and binge drinking; consenters also reported a greater number and more severe psychosocial and mental health problems than those who declined study participation. Taken together, the present study's sample is less

representative of not only general primary care populations but also less representative of the overall clinical population as well.

Second, the extent to which the instruments would perform equally well when delivered by primary care staff and entered into the patient's medical record is not known. It should also be noted that methodological differences including setting, interview mode, and contextual effects make comparisons difficult between other healthcare settings.

Third, while self-administered help promote disclosure of substance use, this format may not be feasible in all practice settings. Self-administration on a tablet could be problematic in patients with low literacy or poor vision, though the audio guidance can help to address these barriers. Elderly patients may also have difficulty using a tablet. Further, tablet computers in primary care settings would require considerations for workflow, security, and hygiene.

Lastly, despite efforts to mask the primary purpose of the screener, it relied solely on self-report information. Also, this study relied on secondary analysis of existing data and was limited by available items. Additionally, participants completed the computerized health survey prior to being offered research participation and there was a consistent order effect, which may have resulted in practical issues (e.g., cumulative time for research at one visit) and may have influenced response rates.

Future Directions: The present study expanded on the current research supporting screening and assessment for substance use in primary care settings. However, further examination of the stability and consistency of self-report of substance use is important. Future research can further evaluate different versions of standardized measures across different levels of data protection and confirm their accuracy in identifying unhealthy alcohol and other drug use in various populations. Future research should also focus on the optimal screening interval and

assessment window for detecting unhealthy alcohol and drug use. The accuracy of screening tools for detecting nonmedical use of prescription drugs, including opioids, should also be expanded upon. Going forward, research on how to best integrate multiple substance components into primary care-based screening assessments and interventions for problem substance use is also needed.

More evidence on important clinical outcomes is also needed, such as longer-term morbidity, mortality, health care utilization, and social and legal outcomes. Further, trials designed to report subgroup effects in diverse populations (e.g., by age, race/ethnicity, sexual identity, or baseline severity) will be important. In particular, research focused on gender-specific screening will be advantageous, especially since research has demonstrated women have increased vulnerability for adverse medical and social consequences associated with substance use (e.g., Polak et al., 2015). Research also suggests that women are more vulnerable to social desirability bias than men (Delker et al., 2016; Durant et al., 2002), as women may be more concerned about judgment or about reporting substance use in a face-to-face interviews due to fear of losing custody of their children. Also, continued research on comorbid mental health and substance abuse disorders will be beneficial. A recent study secondary analysis of the same sample as the present study found that patients who consented to the RCT reported a greater number and more severe psychosocial and mental health problems than those who declined study participation, which has important implications regarding generalizability to other clinic samples (Kelpin et al., 2018).

Future research should also focus on differences of substance use rates at follow-up intervals. Historically, collection of comprehensive baseline assessment data from experimental as well as control group participants has been integral to RCT research. Studies that have begun

to dismantle the components of clinical trial research have found that simply participating in a baseline assessment of substance use behaviors can produce positive changes in subsequent quantity and frequency of alcohol and other drug use. This is important as it highlights the importance of more systematically examining the extent to which simple attention to the problem of heavy/problem substance use may be therapeutic. This most parsimonious interpretation can explain much of the existing data and raises the possibility that remarkably simple approaches might be efficacious.

Finally, future research should build on present study findings and continue to replicate these findings in primary care samples and other samples and across additional substance use outcomes. Also, utilizing a test-retest approach, in which the same instrument(s) are used in both shorter and longer over-lapping timeframes to determine whether the pattern of use is shown to be similar across measures will be important. This will be especially helpful to support their reliability for longitudinal examination of self-reported alcohol and other drug use.

Moreover, while computerized versions of TLFB measures have been developed (Rueger et al., 2012; Sobell et al., 1996) little research exists on how these measures compare to interviewer-administered TLFB or other ACASI procedures; future studies in this area would offer useful information. Further, the comparative performance of the TLFB and ACASI to measure more complex aspects of substance use could be studied.

Conclusions

To our knowledge, this study represents the first examination of computerized anonymous and confidential surveys to a confidential, interviewer-administered assessment across alcohol and illicit drug use as well as prescription drug misuse. Having information on the most parsimonious methods to collect the most sensitive and accurate data possible about

substance use and related behaviors is essential for ensuring the quality and safety of medical care and has the potential to improve medical and psychosocial outcomes. This study supports the use of an anonymous computerized health survey in screening primary care patients for problem substance use. It may also detect alcohol and other drug use problems. However, additional assessment via confidential, interviewer-administered TLFB is needed to capture specific details of use. With advancements in technology, incorporating a confidential, computerized version of the TLFB has the potential to ease barriers and improve cost-effectiveness to incorporating substance screening into busy clinical environments. The United States Preventive Services Task Force (USPSTF) recommends screening for unhealthy alcohol use and recently has recommended screening for unhealthy drug use, in primary care settings in adults 18 years or older. Continued focus on best practices to identify substance use and related problems should remain a core clinical quality measure for primary care settings.

References

Acevedo, A., Panas, L., Garnick, D., Acevedo-Garcia, D., Miles, J., Ritter, G., & Campbell, K. 2018. Disparities in the treatment of substance use disorders: Does where you live matter? *The Journal of Behavioral Health Services & Research,* 45(4), 533-549.

Agerwala, S. M. & McCance-Katz, E. F. (2012). Integrating screening, brief intervention, and referral to treatment (SBIRT) into clinical practice settings: a brief review. *Journal of Psychoactive Drugs*, 44, 307-17.

Ahmad, F. B., Rossen, L. M., Spencer, M. R., Warner, M., & Sutton, P. (2018). Provisional drug overdose death counts. National Center for Health Statistics.

Aikens, J. E. & Rouse, M. E. (2005). Help-seeking for insomnia among adult patients in primary care. *Journal of the American Board of Family Medicine,* 18(4), 257-261.

Ali, R., Awwad, E., Babor, T. F., Bradley, F., Butau, T., & Farrell, M. (2002). The alcohol, smoking and substance involvement screening test (ASSIST): development, reliability and feasibility. *Addiction*, 97(9), 1183-1194.

Ali, R., Meena, S., Eastwood, B., Richards, I., & Marsden, J. (2013). Ultra-rapid screening for substance-use disorders: The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST-Lite). *Drug and Alcohol Dependence*, 132(1-2), 352-361.

Allen, J. P. & Wilson, V. B. (2003). *Assessing alcohol problems: A guide for clinicians and researchers*. Bethesda, MD: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism.

Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gøtzsche, P.C., & Lang, T. (2001). The revised CONSORT Statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine,* 134, 663-694.

Álvarez-Bueno, C., Rodríguez-Martín, B., García-Ortiz, L., Gómez-Marcos, M. A., Martínez-Vizcaíno, V. (2015). Effectiveness of brief interventions in primary health care settings to decrease alcohol consumption by adult non-dependent drinkers: a systematic review of systematic reviews. *Preventive Medicine*, 76, 533-538.

Alter, A. & Yeager, C. (2020). The consequences of COVID-19 on the overdose epidemic: Overdoses are increasing. Overdose Detection Mapping Application Program. Retrieved from: https://files.constantcontact.com/a923b952701/dbf0b5a5-f730-4a6f-a786-47097f1eea78.pdf

American Public Health Association (APHA). (2013). Policy Statement: Defining and Implementing a Public Health Response to Drug Use and Misuse.

American Society of Addiction medicine (ASAM). (2017). Appropriate use of drug testing in clinical addiction medicine. *Journal of Addiction Medicine*, 11, 1-56. doi: 10.1097/ADM.0000000000000322

American Society of Addiction Medicine (2019). Definition of Addiction. Retrieved from: https://www.asam.org/docs/default-source/quality-science/asam's-2019-definition-of-addiction-(1).pdf?sfvrsn=b8b64fc2_2

Andreae, M. H., Rhodes, E., Bourgoise, T., Carter, G. M., White, R. S., Indyk, D., ... Rhodes, R. (2016). An ethical exploration of barriers to research on controlled drugs. The American Journal of Bioethics, 16(4), 36-47.

Andrews, G. & Slade, T. (2001). Interpreting scores on the Kessler Psychological Distress Scale (k10). *Australian and New Zealand Journal of Public Health*, 25, 494-497.

Appel, J., von der Pütten, A., Krämer, N. C., & Gratch, J. (2012). Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Advances in Human-Computer Interaction*, Article ID 324694.

Babor, T. F., Brown, J., & Del Boca, F. K. (1990). Validity of self-reports in applied-research on addictive behaviors - Fact or fiction. *Behavioral Assessment*, 12(1), 5-31.

Babor. T. & Higgins-Biddle, J. (2001). Brief interventions for hazardous and harmful drinking: A manual for use in primary care. World Health Organization, Department of Mental Health and Substance Dependence.

Babor, T. F., & Kadden, R. M. (2005). Screening and interventions for alcohol and drug problems in medical settings: what works? *Journal of Trauma and Acute Care Surgery*, 59(3), S80-S87.

Babor, T. F., Del Boca, F., & Bray, J. W. (2017). Screening, brief intervention and referral to treatment: implications of SAMHSA's SBIRT initiative for substance abuse policy and practice. *Addiction*, 112, 110-117.

Babor, T. F., Higgins-Biddle, J. C., Saunders, J., and Monteiro, M. G. (2001). AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Health Care, Second Edition: World Health Organization.

Babor, T. F., McRee, B. G., Kassebaum, P. A., Grimaldi, P. L., Ahmed, K, & Bray, J. (2007). Screening, Brief Intervention, and Referral to Treatment (SBIRT): toward a public health approach to the management of substance abuse. *Substance Abuse*, 28(3), 7-30.

Babor, T. F., Steinberg, K., Anton, R., & Del Boca, F. (2000). Talk is cheap: Measuring drinking outcomes in clinical trials. *Journal of Studies on Alcohol and Drugs*, 61(1), 55-63.

Baggio, S., Iglesias, K., Studer, J., Dupuis, M., Daeppen, J. B., & Gmel, G. (2015). Is the relationship between major depressive disorder and self-reported alcohol use disorder an artificial one? *Alcohol and Alcoholism*, 50, 195-199.

Bahorik, A. L., Satre, D. D., Kline-Simon, A. H., Weisner, C. M., & Campbell, C. I. (2017). Alcohol, cannabis, and opioid use disorders, and disease burden in an integrated health care system. *Journal of Addiction mMdicine*, 11(1), 3-9.

Barbosa, C., Cowell, A., Bray, J., & Aldridge, A. (2015). The cost-effectiveness of alcohol Screening, Brief Intervention, and Referral to Treatment (SBIRT) in emergency and outpatient medical settings. *Journal of Substance Abuse Treatment*, 53, 1-8.

Bastien, C. H., Vallières, A., & Morin, C. M. (2001). Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Medicine*, 2, 297-307.

Basurto, F. Z., Montes, J. G., Cubos, P. F., Santed, F. S., Ríos, F. L., & Moreno, A. M. (2009). Validity of the self-report on drug use by university students: Correspondence between self-reported use and use detected in urine. *Psicothema*, 21(2), 213-219.

Bazzi, A & Saitz, R. (2018). Screening for unhealthy alcohol use. *Journal of the American Medical Association*, 320(18), 1869-1871.

Beatty, J. R., Chase, S. K., & Ondersma, S. J. (2014). A randomized study of the effect of anonymity, quasi-anonymity, and Certificates of Confidentiality on postpartum women's disclosure of sensitive information. *Drug and Alcohol Dependence*, 134, 180-184.

Beck, F., Guignard, R., & Legleye, S. (2014). Does computer survey technology improve reports on alcohol and illicit drug use in the general population? A comparison between two surveys with different data collection modes in France. *PLoS ONE*, 9(1), e85810.

Berman, A. H., Bergman, H., Palmstierna, T., & Schlyter, F. (2005). Evaluation of the Drug Use Disorders Identification Test (DUDIT) in criminal justice and detoxification settings and in a Swedish population sample. *European Addiction Research*, 11, 22-31.

Berner, M. M., Kriston, L., Bentele, M., & Härter, M. (2007). The alcohol use disorders identification test for detecting at-risk drinking: A systematic review and meta-analysis. *Journal of Studies on Alcohol and Drugs*, 68, 461-473.

Bernstein, J., Bernstein, E., Tassiopoulos, K., Heeren, T., Levenson, S., & Hingson, R. (2005). Brief motivational intervention at a clinic visit reduces cocaine and heroin use. *Drug Alcohol Dependence*, 77(1), 49-59.

Bing, E. G., Burnam, M. A., Longshore, D., Fleishman, J. A., Sherbourne, C. D., London, A. S., ... Shapiro, M. (2001). Psychiatric disorders and drug use among Human Immunodeficiency Virus-infected adults in the United States. *Archives of General Psychiatry*, 58(8), 721-728.

Bjarnason, T. & Adalbjarnardottir, S. (2000). Anonymity and confidentiality in school surveys on alcohol, tobacco, and cannabis use. *Journal of Drug Issues,* 30, 335-343.

Bobak, M., Room, R., Pikhart, H., Kubinova R., Malyutina, S., Pajak A., Kurilovitch, S., ...Marmot, M. (2004). Contribution of drinking patterns to differences in rates of alcohol related problems between three urban populations. *Journal of Epidemiology and Community Health*, 58(3), 238-242.

Bogenschutz, M. P., Donovan, D. M., Mandler, R. N., Perl, H. I., Forcehimes, A. A., Crandall, C., … Douaihy, A. (2014). Brief Intervention for Drug Users Presenting in Emergency Departments (NIDA CTN Protocol 0047: SMART-ED). *JAMA Internal Medicine*, 174(11), 1736-1745.

Bone, C., Gelberg, L., Vahidi, M., Leake, B., Yacenda-Murphy, J., Andersen, R. M. (2016). Under-reporting of Risky Drug Use Among Primary Care Patients in Federally Qualified Health Centers. *Journal of Addiction Meicine,*10(6), 387-394.

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, 27(3), 281-291.

Bradley, K. A., Bush, K. R., Epler, A. J., Dobie, D. J., Davis, T. M., Sporleder, J. L., Maynard C., ... Kiviahan, D. R. (2003). Two brief alcohol-screening tests From the Alcohol Use Disorders Identification Test (AUDIT): validation in a female Veterans Affairs patient population. *Archives of Internal Medicines,* 163(7), 821-829.

Bradley, K. A., Lapham, G. T., Hawkins, E. J., Achtmeyer, C. E., Williams, E. C., Thomas, R. M., & Kivlahan, D. R. (2011). Quality concerns with routine alcohol screening in VA clinical settings. *Journal of General Internal Medicine,* 26(3), 299-306.

Bray, J. W., Mallonee, E., Dowd, W., Aldridge, A., Cowell, A. J., & Vendetti, J. (2014). Program-and service-level costs of seven screening, brief intervention, and referral to treatment programs. *Substance Abuse and Rehabilitation*, 5, 63-73.

Bray, J. W., Zarkin, G. A., Hinde, J. M., & Mills, M. J. (2012). Costs of alcohol screening and brief intervention in medical settings: a review of the literature. *Journal of Studies on Alcohol and Drugs,* 73(6), 911-919.

Brandimarte, L., Acquisti, A., & Loewenstein, G. (2012). Misplace confidences: Privacy and the control paradox. *Social Psychological and Personality Science*, 4(3), 350-347.

Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies,* 62(2), 161-178.

Brown, R. L., & Rounds, L. A. (1995). Conjoint screening questionnaires for alcohol and other drug abuse: Criterion validity in a primary care practice. *Wisconsin Medical Society*, 94, 135-140.

Buchsbaum, D.G., Welsh, J., Buchanan, R. G., & Elswick, R. K. (1995). Screening for drinking problems by patient self-report: Even 'safe' levels may indicate a problem. *Archives of Internal Medicine,* 155(1), 104-108.

Burns, E., Gray, R., & Smith, L. A. (2010). Brief screening questionnaires to identify problem drinking during pregnancy: a systematic review. *Addiction,* 105(4), 601-614.

Bush, K., Kivlahan, D. R., McDonell, M. B., Fihn, S. D., & Bradley, K. A. (1998). The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. *Archives of Internal Medicine,* 158(16), 1789-1795.

Butler, S. F., Villapiano, A., & Malinow, A (2009). The effect of computer-mediated administration on self-disclosure of problems on the addiction severity index. *Journal of Addiction Medicine,* 3(4), 194-203.

Butler, S. F., Chiauzzi, E., Bromberg, J. I., Budman, S. H., Buono, D. P. (2003). Computer-assisted screening and intervention for alcohol problems in primary care. *Journal of Technology in Human Services*, 21(3), 1-19.

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Research*, 28, 193-213.

Caetano, R., Schafer, J., & Cunradi, C. B. (2001). Alcohol-related intimate partner violence among white, black, and Hispanic couples in the United States. *Alcohol Research & Health*, 25, 58-65.

Carey, K. B., Carey, M. P., Maisto, S. A., & Henson, J. M. (2004). Temporal stability of the Timeline Followback interview for alcohol and drug use with psychiatric outpatients. *Journal of Studies on Alcohol,* 65, 774-781.

Carliner, H., Delker, E., Fink, D. S., Keyes, K. M., & Hasin, D. S. (2016). Racial discrimination, socioeconomic position, and illicit drug use among US Blacks. *Social Psychiatry and Psychiatric Epidemiology*, 51(4), 551-560.

Center for Behavioral Health Statistics and Quality (2016). Results from the 2015 National Survey on Drug Use and Health: Detailed tables. Substance Abuse and Mental Health Services Administration.

Center for Substance Abuse Treatment. (1997). A guide to substance abuse services for primary care clinicians.

Chermack, S. T., Roll, J., Reilly, M., Davis, L., Kilaru, U., & Grabowski, J. (2000). Comparison of patient self-reports and urinalysis results obtained under naturalistic methadone treatment conditions. *Drug and Alcohol Dependence*, 59, 43-49.

Clark, W. B., & Midanik, L.T. (1982). Alcohol Use and Alcohol Problems Among U.S. Adults: Results of the 1979 Survey. *Alcohol and Health Monograph No. 1.* National Institute on Alcohol Abuse and Alcoholism.

Clark, C. B., Zyambo, C. M., Li, Y., & Cropsey, K. L. (2016). The impact of non-concordant self-report of substance use in clinical trials research. *Addictive Behaviors*, 58, 74-79.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.

Cohen, J. (1998). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum Associates.

Cohen, E., Feinn, R., Arias, A., & Kranzler, H. R. (2007). Alcohol treatment utilization: Findings from the National Epidemiologic Survey on Alcohol and Related Conditions. *Drug and Alcohol Dependence*, 86, 214-221.

Colón, H. M., Robles, R. R., & Sahai, H. (2002). The validity of drug use self-reports among hard core drug users in a household survey in Puerto Rico: Comparison of survey responses of cocaine and heroin use with hair tests. *Drug and Alcohol Dependence*, 67(3), 269-279.

Compton, W. M., Jones, C. M., & Baldwin, G. T. (2016). Relationship between Nonmedical Prescription-Opioid Use and Heroin Use. *New England Journal of Medicine*, 374(2), 154-63.

Cone, E. J. (1993). Saliva testing for drugs of abuse. *Annals of the New York Academy of Sciences*, 694(1), 91-127.

Cone, E. J. (1997). New developments in biological measures of drug prevalence. *NIDA Research Monograph Index*, 167, 108-29.

Cone, E. J. (2012). Oral fluid results compared to self reports of recent cocaine and heroin use by methadone maintenance patients. *Forensic Science International*, 215(1), 88-91.

Connors, G. J. and Volk, R. J. (2004). Self-report screening for alcohol problems among adults. Accessed https://pubs.niaaa.nih.gov/publications/assessingalcohol/selfreport.htm

Copeland, J. (1997). A qualitative study of barriers to formal treatment among women who self-managed change in addictive behaviours. *Journal of Substance Abuse Treatment,* 14(2), 183-90.

Corrigan, P. W., Schomerus, G., Shuman, V., Kraus, D., Perlick, D., Harnish, A., Kulesza, M., ... & Smelson, D. 2017. Developing a research agenda for reducing the stigma of addictions, part II: Lessons from the mental health stigma literature. The *American Journal on Addictions,* 26(1), 67-74.

Cottler, L. B. (2000). Department of Psychiatry. Washington University School of Medicine. Composite International Diagnostic Interview - Substance Abuse Module (SAM).

Cottler, L. B., Robins, L. N., Helzer, J. E. (1989). The reliability of the CIDI-SAM: a comprehensive substance abuse interview. *British Journal of Addiction to Alcohol and Other Drugs,* 84(7), 801-814.

Cowell, A.J., Dowd, W.N., Mills, M.J., Hinde, J.M., & Bray, J.W. (2017). Sustaining SBIRT in the wild: simulating revenues and costs for Screening, Brief Intervention and Referral to Treatment programs. *Research Report,* 112(S2), 101-109.

Cuijpers P., Riper H., and Lemmers L. (2004). The effects on mortality of brief interventions for problem drinking: a meta-analysis. *Addiction,* 99, 839-845.

Cyr, M. G. & Wartman, S. A. (1988). The effectiveness of routine screening questions in the detection of alcoholism. *Journal of the American Medical Association,* 259(1), 51-54.

Czeisler, M. É., Petrosky, L. R., Wiley, J. F., Christensen, A., Njai, R., Weaver, M. D., Robbines, R., ... Rajartnam, S. M. (2020). Mental health, substance use, and suicidal ideation during the COVID-19 Pandemic – United States, June 24-30, 2020. *Morbidity and Mortality Weekly Report (MMWR),* 69, 1049-1057.

Darke, S. (1998). Self-report among injecting drug users: a review. *Drug and Alcohol Dependence,* 51, 253-263.

Davenport, S., Gray, T. J., & Melek, S. P. (2020). How do individuals with behavioral health conditions contribute to physical and total healthcare spending? Milliman Research Report.

Davis, C. G., Thake, J., & Vilhena, N. (2010). Social desirability biases in self-reported alcohol consumption and harms. *Addicted Behaviors,* 35(4), 302-311.

Dawson, D. A., Smith, S. M., Saha, T. D., Rubinsky, A. D., & Grant, B. F. (2012). Comparative performance of the AUDIT-C in screening for DSM-IV and DSM-5 alcohol use disorders. *Drug and Alcohol Dependence,* 126, 384-8.

de Meneses-Gaya, C., Zuardi, A. W., Loureiro, S. R., & Crippa, A. S. J. (2009). Alcohol use disorders identification test (AUDIT): an updated systematic review of psychometric properties. *Psychology & Neuroscience,* 2(1), 83-97.

Del Boca, F. K & Darkes. J. (2003). The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction,* 98(2), 1-12.

Delker, E., Aharonovich, E., & Hasin, D. (2016). Interviewer-administered TLFB vs. self-administered computerized (A-CASI) drug use frequency questions: a comparison in HIV-infected drug users. *Drug and Alcohol Dependence,* 161, 29-35.

Dennis, M. L., Funk, R., Godley, S. H., Godley, M. D., & Waldron, H. (2004). Cross-validation of the alcohol and cannabis use measures in the Global Appraisal of Individual Needs (GAIN) and Timeline Followback (TLFB; Form 90) among adolescents in substance abuse treatment. *Addiction*, 99(Suppl 2), 120-128.

Dhalla, S. & Kopec, J. A. (2007). The CAGE questionnaire for unhealthy alcohol use: A review of reliability and validity studies. *Clinical & Investigative Medicine*, 30(1), 33-42.

Digiusto, E. & Treloar, C. (2007). Equity of access to treatment, and barriers to treatment for illicit drug use in Australia. *Addiction*, 102(6), 958-969.

Dolezal, C., Marhefka, S. L., Santamaria, E. K., Leu, C. S., Brackis-Cott, E., & Mellins, C. A. (2012). A comparison of audio computer-assisted self-interviews to face-to-face interviews of sexual behavior among perinatally HIV-exposed youth. *Archives of Sexual Behavior*, 41, 401-410.

Donovan, D. M., Bigelow, G. E., Brigham, G. S., Carroll, K. M., Cohen, A. J., Gardin, J. G., ... Wells, E.A. (2012). Primary outcome indices in illicit drug dependence treatment research: systematic approach to selection and measurement of drug use end-points in clinical trials. *Addiction*, 107, 694-708.

Durant, L. E., Carey, M. P., & Schroder, K. E. E. (2002). Effects of anonymity, gender, and erotophilia on the quality of data obtained from self-reports of socially sensitive behaviors. *Journal of Behavioral Medicine*, *2*5(5), 438-467.

Edelman, E. J., & Tetrault, J. M. (2019). Unhealthy alcohol use in primary care: the elephant in the examination room. *JAMA Intern Med*icine, 179(1), 9-10.

Edlund, M. J., Booth, B. M., & Feldman, Z. L. (2009). Perceived need for treatment for alcohol use disorders: Results from two national surveys. *Psychiatric Services*, 60, 1618-1628.

Edlund, M. J., Unützer, J., & Curran, G. M. (2006). Perceived need for alcohol, drug, and mental health treatment. *Social Psychiatry and Psychiatric Epidemiology*, 41, 480-487.

Edlund, M. J., Unutzer, J., & Wells, K. B. (2004). Clinician screening and treatment of alcohol, drug, and mental problems in primary care: results from healthcare for communities. *Medical Care*, 42(12), 1158-66.

Ehrman, R. N. & Robbins, S. J. (1994). Reliability and validity of 6-month timeline reports of cocaine and heroin use in a methadonepopulation. *Journal of Consulting and Clinical Psychology,* 62, 843-850.

Esser, M. B., Sacks, J. J., Sherk, A., Karriker-Jaffe, K. J., Greenfield, T. K., Pierannunzi, C., & Brewer, R. D. (2020) (in press). Distribution of drinks consumed by U.S. adults by average daily alcohol consumption: A comparison of 2 nationwide surveys. *American Journal of Preventative Medicine,* 1-9.

Esser, M. B., Sherk, A., Liu, Y., Nami, T. S., Stockwell, T., Stahre, M., Kanny, D., ... Brewer, R. D. (2020). Deaths and Years of Potential Life Lost From Excessive Alcohol Use — United States, 2011–2015. *Morbidity and Mortality Weekly Report* (*MMWR*), 69, 1428-1433.

Estes, L. J., Lloyd, L. E., Teti, M., Raja, S., Bowleg, L., Allgood, K. L., & Glick, N. (2010). Perceptions of audio computer-assisted self-interviewing (ACASI) among women in an HIV-positive prevention program. *PLoS One*, 5:e9149.

Ewing, J. A. (1984). Detecting alcoholism: The CAGE questionnaire. *Journal of the American Medical Association,* 252, 1905-1907.

Falk, D., Yi, H. & Hiller-Sturmhöfel, S. (2008). An epidemiologic analysis of co-occurring alcohol and drug use and disorders: Findings from the National Epidemiologic Survey of Alcohol and Related Conditions (NESARC). *Alcohol Research & Health,* 31(2), 100-110.

Fals-Stewart, W., O'Farrell, T. J., Freitas, T. T., McFarlin, S. K., & Rutigliano, P. (2000). The timeline followback reports of psychoactive substance use by drug-abusing patients: psychometric properties. *Journal of Consulting and Clinical Psychology*, 68(1), 134-44.

Feldhaus, K. M., Kozoil-McLain, J., Amsbury, H. L., Norton, I. M., Lowenstein, S. R., & Abbott J. T. (1997). Accuracy of 3 brief screening questions for detecting partner violence in the emergency department. *Journal of the American Medical Association,* 277(17), 1357-1361.

Fendrich, M., Johnson, T. P., Wislar, J. S., Hubbell, A., & Spiehler, V. (2003). The utility of drug testing in epidemiological research: results from a general population survey. *Addiction,* 99(2), 197-208

Fiellin, D. A., Reid, M. C., & O'Connor, P. G. (2000). Screening for alcohol problems in primary care: a systematic review. *Archives of Internal Medicine*, 160, 1977-1989.

Fiellin, D. A., McGinnis, K. A., Maisto, S. A., Justice, A. C., & Bryant, K. (2013). Measuring Alcohol Consumption Using Timeline Followback in Non-Treatment-Seeking Medical Clinic Patients With and Without HIV Infection: 7-, 14-, or 30-day Recall. *Journal of Studies on Alcohol and Drugs*, 74(3), 500-504.

Flegal, K. M. (1990). Agreement between two dietary methods in the measurement of alcohol consumption. *Journal of Studies on Alcohol and Drugs*, 51, 408-414.

Fleming, M. F. & Barry, K. L. (1991). A three-sample test of a masked alcohol screening questionnaire. *Alcohol Alcohol*, 26, 81-91.

Fleming, M. F., Balousek, S. L., Kelssig, C. L., Mundt, M. P., & Brown, D. D. (2007). Substance use disorders in a primary care sample receiving daily opioid therapy. *The Journal of Pain*, 8(7), 573-582.

Freeman, E., McGuire, C. A., Thomas, J. W., & Thayer, D. A. (2014). Factors affecting costs in Medicaid populations with behavioral health disorders. *Medical Care*, 52(3), S60-S66.

Gerteis, J., Izrael, D., Deitz, D., LeRoy, L., Ricciardi, R., Miller, T., & Basu J. (2014). Multiple Chronic Conditions Chartbook. Agency for Healthcare Research and Quality: AHRQ Publications No, Q14-0038.

Geurero, E. G., Marsh, J. C., Duan, L., Oh, C., Perron, B., & Lee, B. (2013). Disparities in completion of substance abuse treatment between and within racial and ethnic groups. *Health Services Research*, 48(4), 1450-1467.

Ghitza, U. E., Gore-Langton, R. E., Lindblad, R., Shide, D., Subramaniam, G., & Tai, B. (2013). Common data elements for substance use disorders in electronic health records: the NIDA Clinical Trials Network experience. *Addiction*, 108, 3-8.

Gnambs, T. & Kaspar, K. (2015). Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. *Behavior research methods*, 47(4), 1237-1259.

Gomes, T., Tadrous, M., Muhammad, M., Paterson, M.J., & Juurlink, D.M. (2018). The Burden of Opioid-Related Mortality in the United States. *JAMA Network Open*, 1(2):e180217.

Gorfinkel, L. R., Stohl, M., & Hasn, D. (2020). Association of depression with past-month cannabis use among US adults aged 30 to 59 years, 2005 to 2016. *JAMA Network Open*, 3(8): e2013802.

Graham, K., Cheng, J., Bernards, S., Wells, S., Rehm, & Kurdyak, P. (2017). How much do mental health and substance use/addiction affect use of general medical services? Extent of use, reason for use, and associated costs. *Canadian Journal of Psychiatry*, 62(1), 48-56.

Grant, B. F. (1997). Barriers to alcoholism treatment: Reasons for not seeking treatment in a general population sample. *Journal of Studies on Alcohol*, 58, 365-371.

Grant, B.F., Saha, T. D., Ruan, W. J., Goldstein, R. B., Chou, S. P., Jung, J., Zhang, H., ... Hasin, D. S. (2016). Epidemiology of DSM-5 drug use disorder: results from the national epidemiologic survey on alcohol and related conditions-III. *JAMA Psychiatry*, 73(1), 39-47.

Grant, B. F., Chou, S. P., Tulshi, T. D., Pickering, R. P., Kerridge, B. T., Ruan, R. W., ... Hasin, D.S. (2017). Prevalence of 12-Month Alcohol Use, High-Risk Drinking, and DSM-IV Alcohol Use Disorder in the United States, 2001-2002 to 2012-2013 Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Jama Psychiatry*, 74(9), 911-923.

Grove, W. M., Andreasen, N. C., McDonald-Scott, P., Keller, M. B., & Shapiro, R. W. (1981). Reliability studies of psychiatric diagnosis: theory and practice. *Archives of General Psychiatry*, 38, 408-413.

Grekin, E. R., Svikis, D. S., Lam, P., Connors, V., Lebreton, J. M., Streiner, D. L., Smith, C., & Ondersma, S. J. (2010). Drug use during pregnancy: validating the Drug Abuse Screening Test against physiological measures. *Psychology of Addictive Behaviors*. 24(4), 719-723.

Grucza, R. A., Sher, K. J., Kerr, W. C., Kruass, M. J., Lui, C. K., McDowell, Y. E., Hartz, S., ... & Bierut, L. J., (2018). Trends in adult alcohol use and binge drinking in the early 21st-Century United States: A meta-analysis of 6 national survey series. *Alcoholism: Clinical and Experimental Research*, 42(10), 1939-1950.

Gryczynski, J., McNeely, J., Wu, L. T., Subramaniam, G. A., Svikis, D. S., Cathers, L. A., Sharma, G., ... Schwartz, R. P. (2017). Validation of the TAPS-1: A Four-Item Screening Tool to Identify Unhealthy Substance Use in Primary Care. *Journal of General Internal Medicine*, *32*(9), 990-996.

Gupman, A. E., Svikis, D., McCaul, M. E., Anderson, J., & Santora, P. B. (2002). Detection of alcohol and drug problems in an urban gynecology clinic. *Journal of Reproductive Medicine*, 47(5), 404-410.

Hall, W. & Farrell, M. (1997). Comorbidity of mental disorders with substance misuse. *British Journal of Psychiatry*, 171(7), 4-5.

Halkitis, P. N., Solomon, T. M., Moeller, R. W., Doig, S. A. R., Espinosa, L. S., Siconolfi, D., & Homer, B. D. (2009). Methamphetamine use among gay, bisexual and non-identified men-who-have-sex-with-men: An analysis of daily patterns. *Journal of Health Psychology*, 14(2), 222-231.

Hamid, R., Deren, S., Beardsley, M., & Tortu, S. (1999). Agreement between urinalysis and self-reported drug use. *Substance Use and Misuse*, 34(11), 1585-1592.

Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36.

Hargraves, D., White, C., Frederick, R., Cinibulk, M, Peters, M., Young, A, & Elder, N. (2017). Implementing SBIRT (Screening, Brief Intervention and Referral to Treatment) in primary care: lessons learned from a multi-practice evaluation portfolio. *Public Health Review,* 38**(**31). https://doi.org/10.1186/s40985-017-0077-0

Harrison, L. (1997). The validity of self-reported drug use in survey research: An overview and critique of research methods. In: Harrison, L. & Hughes, A. (Eds.). The validity of self-reported drug use: Improving the accuracy of survey estimates. U.S. Department of Health and Human Services, National Institute on Drug Abuse, pp. 17–36.

Harrison, L. & Hughes, H. (1997). NIDA Research Monograph, No 167. Department of Health and Human Services; 1997. The validity of self-reported drug use: Improving the accuracy of survey estimates; pp. 1-16.

Hasin D. S. (2018). US Epidemiology of Cannabis Use and Associated Problems. Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology, 43(1), 195-212.

Haass-Koffler, C. L., Piacentino, D., Li, X., Long, V. M., Lee, M. R., Swift, R. M., Kenna, G. A., & Leggio, L. (2020). Differences in sociodemographic and alcohol-related clinical characteristics between treatment seekers and nontreatment seekers and their role in predicting outcomes in the COMBINE Study of Alcohol Use Disorder. *Alcoholism: Clinical and Experimental Research*. doi: 10.1111/acer.14428. Epub ahead of print. PMID: 32997422.

Hayes, S. L., McCarthy, D., & Radley, D. (2016). The impact of a behavioral health condition on high-need adults. To the point. The Commonwealth Fund; New York, NY. Retrieved from https://www.commonwealthfund.org/publications/issue-briefs/2016/aug/high-need-high-cost-patients-who-are-they-and-how-do-they-use

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerstrom, K. O. (1991). The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *British Journal of Addition to Alcohol and Other Drugs*, 86(9), 1119-1127.

Hester, R. K. & Squires, D. D. (2004). Outcome research: alcoholism. In Galanter, M., Kleber H.D., (Eds.), *Textbook of Substance Abuse Treatment* (p. 129-135). American Psychiatric Association.

Hettema, J., Steele, J., & Miller, W. (2005). Motivational Interviewing. *Annual Review of Clinical Psychology*, 91-111.

Higgins, R. & Compton, W. M. (2014). Screening and brief intervention and referral to treatment for drug use in primary care: back to the drawing board. *Journal of the American Medical Association,* 312(5), 488-489.

Hildebrand, M. (2015). The psychometric properties of the drug use disorders identification test (DUDIT): a review of recent research. *Journal of Substance Abuse Treatment*, 53, 52–59.

Hirsh, A. T., Anastas, T. M., Miller, M. M., Quinn, P. D., & Kroenke, K. (2020). Patient race and opioid misuse history influence provider risk perceptions for future opioid-related problems. *American Psychologist, 75*(6), 784-795.

Hjorthøj, C. R., Hjorthøj, A. R., & Nordentoft, M. (2012). Validity of Timeline Follow-Back for self-reported use of cannabis and other illicit substances--systematic review and meta-analysis. *Addicted Behaviors*, 37, 225-233.

Hoeppner, B. B., Stout, R. L., Jackson, K. M., & Barnett, N. P. (2010). How good is fine-grained timeline follow-back data? Comparing 30-day tlfb and repeated 7-day tlfb alcohol consumption reports on the person and daily level. *Addictive Behaviors*, 35(12), 1138-1143.

Hoggatt, K. J., Frayne, S. M., Saechao, F. S., Yano, E. M., & Washington, D. L. (2019) Substance use disorder-related disparities in patient experiences of primary care. *Health Equity* 3(1), 193-197.

Hormes, J. M., Gerhardstein, K. R., & Griffin, P. T. (2012). Under-Reporting of Alcohol and Substance Use versus other Psychiatric Symptoms in Individuals living with HIV. *AIDS Care*, 24(4), 420-423.

Humeniuk, R., Ali, R., Babor, T., Souza-Formigoni, M. L., de Lacerda, R. B., Ling, W., ... Vendetti, J. (2012). A randomized controlled trial of a brief intervention for illicit drugs linked to the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) in clients recruited from primary health-care settings in four countries. *Addiction*, 107, 957-66.

Humeniuk, R., Ali, R., Babor, T.F., Farrell, M., Formigoni, M.L., Jittiwutikarn, J., ... Simon, S. (2008). Validation of the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST). *Addiction,*103(6), 1039-47.

IBM Corp. Released 2020. IBM SPSS Statistics for Macintosh, Version 27.0. Armonk, NY: IBM Corp.

Islam, M. M., Topp, L., Conigrave, K. M., van Beek, I., Maher, L., White, A., Rodgers, C., & Day, C. A. (2012). The reliability of sensitive information provided by injecting drug users in a clinical setting: clinician-administered versus audio computer-assisted self-interviewing (ACASI). *AIDS Care,* 24(12), 1496-1503.

Jain, R., Tripathi, B., Varghese, S., Kumar, N., & Kumar, M. 2006. Can you trust self-reports among injectable drug users in clinical setting? *Journal of Substance Use,* 11(6):431-436.

Jastrzębska, I., Zwolak, A., Szczyrek, M., Wawryniuk, A., Skrzydło-Radomańska, B., & Daniluk, J. (2016). Biomarkers of unhealthy alcohol use: recent advances and future prospects. *Przegląd Gastroenterologiczny*, 11(2), 78-89.

John, W. S, Zhu, H., Mannelli, P., Schwartz, R. P., Subramaniam, G. A., and Wu, L. T. (2018). Prevalence, patterns, and correlates of multiple substance use disorders among adult primary care patients. *Drug and Alcohol Dependence*, 187(1), 79-87.

Johnson, T. & Fendrich, M. (2005). Modeling sources of self-report bias in a survey of drug use epidemiology. *Annals of Epidemiology*, 15, 381-389.

Johnson, N. B., Hayes, L. D., Brown, K., Hoo, E. C., & Ethier, K. A. (2014). Centers for Disease Control and Prevention (CDC) National Health Report: Leading causes of morbidity and mortality and associated behavioral risk and protective factors—United States, 2005–2013. *Morbidity and Mortality Weekly Report (MMWR*), 63(Suppl), 3-27.

Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology*, 31, 177-192.

Jonas, D. E., Garbutt, J. C., Brown, J. M., Amick, H. R., Brownley, K. A., Council, C. L., ... Harris, R. P. (2012). Screening, behavioral counseling, and referral in primary care to reduce unhealthy alcohol use. Comparative Effectiveness Review No. 64. AHRQ Publication No. 12-EHC055-EF. Agency for Healthcare Research and Quality.

Jones, A. L., Mor, M. K., Cashy, J. P., Gordon, A. J., Haas, G. L., Schaefer, J. H., & Hausmann, L. R. (2016). Racial/Ethnic differences in primary care experiences in Patient-Centered Medical Homes among veterans with mental health and substance use disorders. *Journal of General Internal Medicine*, 31(12), 1435-1443.

Kaner, E. F. S., Beyer, F. R., Muirhead, C, Campbell, F, Pienaar, E. D., Bertholet, N., ... Burnand, B. (2018). Effectiveness of brief alcohol interventions in primary care populations. *Cochrane Database of Systematic Reviews*, Issue 2. Art. No.: CD004148.

Kaner, E. F., Dickinson, H. O., Beyer, F., Pienaar, E., Schlesinger, C., & Campbell, F., ... Heather, N. (2009). The effectiveness of brief alcohol interventions in primary care settings: A systematic review. *Drug and Alcohol Review*, 28(3), 301-323.

Kanny, D., Naimi, T. S., Liu, Y., Lu, H., & Brewer, R. D. (2018). Annual total binge drinks consumed by U.S. adults, 2015. *American Journal of Preventive Medicine*, 54(4), 486-496.

Kanny, D., Naimi, T. S., Liu, Y, & Brewer, R. D. (2020). Trends in total binge drinks per adult who reported binge drinking — United States, 2011–2017. *Morbidity and Mortality Weekly Report (MMWR)*, 69(2), 30-34.

Kasunic, A. & Lee, M. A. (2014). Societal burden of substance abuse. *International Public Health Journal*, 6(3), 269-282.

Kelpin, S. S., Ondersma, S. J., Weaver, M., & Svikis, D. S. (2018). Representativeness of patients enrolled in a primary care clinical trial for heavy/problem substance use. *Substance Abuse, 39*(4), 469-475.

Kelpin, S. S., Rusteikas, S. G., Karjane, N. W., & Svikis, D. S. (2019). Screening for At-Risk Alcohol and Drug Use in the Antenatal Period: How Do Young Women Compare with Older Adult Women? *Journal of Pediatric and Adolescent Gynecology*, 32(3), 325-329.

Kessler, R. & Mroczek, D. (1992). An update of the development of mental health screening scales for the US national health interview study. Survey Research Center of the Institute for Social Research. University of Michigan.

Kessler, R.C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Norman, S. L., Walters, E. E., & Zaslavsky, A. M. (2002) Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32(6), 959-956.

Kessler, R. C., Berglund, P., Demmler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distribution of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 593-602.

Keyes, K. M., Hatzenbuehler, M. L., McLaughlin, K. A., Link, B., Olfson, M., Grant, B. F., & Hasin, D. (2010). Stigma and treatment for alcohol disorders in the United States. *American Journal of Epidemiology*, 172, 1364-1372.

Khan, M. R., Young, K. E., Caniglia, E. C., Fiellin, D. A., Maisto, D. A., Marshall, B. D., Edelman, J., ... Braithwaite, S. R. (2020). Association of alcohol screening scores with adverse mental health conditions and substance use among US adults. *JAMA Network Open*, 3(3), e200895. doi:10.1001/jamanetworkopen.2020.0895.

Kidwella, D.A., Holland, J.C., and Athanaselisc, A. (1998). Testing for drugs of abuse in saliva and sweat. *Journal of Chromatography B: Biomedical Sciences and Applications*, 713(1), 111-135.

Killeen, T. K., Brady, K. T., Gold, P. B., Tyson, C., & Simpson, K. N. (2004). Comparison of self-report versus agency records of service utilization in a community sample of individuals with alcohol use disorders. *Drug Alcohol Dependence*, 73(2), 141-147.

Kim, T. W., Bernstein, J., Cheng, D. M., Lloyd-Travaglini, C., Samet, J. H., Palfai, T. P., & Saitz, R. (2017). Receipt of addiction treatment as a consequence of a brief intervention for drug use in primary care: a randomized trial. *Addiction, 112(5)*, 818-827.

Kim, J. K., Dubowitz, H., Hudson-Martin, E., & Lane W. (2008). Comparison of three data Collection Methods for Gathering Sensitive and Non-sensitive Information. *Ambulatory Pediatrics*, 8, 255-260.

Koo, T. K. & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155-163.

Korthuis, P. T., McCarty, D., Weimer, M., Bougatsos, C., Blazina, I., Zakher, B., Grusing, S., ... & Chou, R., (2017). Primary care-based models for the treatment of opioid use disorder: A scoping review. *Annals of Internal Medicine*, 166(4), 268-278.

Krupski, A., Joesch, J. M., Dunn, C., Donovan, D., Bumgardner, K., Lord, S. P., ... Roy-Byrne, P. (2012). Testing the effects of brief intervention in primary care for problem drug use in a randomized controlled trial: rationale, design, and methods. *Addiction Science and Clinical Practice*, 7(1), 27-37.

Kulesza, M., Larimer, M. E., & Rao, D. (2013). Substance use related stigma: what we know and the way forward. *Addicted Behaviors, Therapy, & Rehabilitation*. 27(2), 2-6.

Kumar, P. C., Cleland, C. M., Gourevitch, M. N., Rotrosen, J., Strauss, S., Russell, L., & McNeely, J. (2016). Accuracy of the Audio Computer Assisted Self Interview version of the Alcohol, Smoking and Substance Involvement Screening Test (ACASI ASSIST) for identifying unhealthy substance use and substance use disorders in primary care patients. *Drug and Alcohol Dependence*, 165, 38-44.

Kypri, K., Langley, J. D., Saunders, J. B., & Cashell-Smith, M. L. (2007). Assessment may conceal therapeutic benefit: findings from a randomized controlled trial for hazardous drinking. *Addiction*, 102(1), 62-70.

Lanier, D. & Ko, S. (2008). Screening in Primary Care Settings for Illicit Drug Use: Assessment of Screening Instruments — A Supplemental Evidence Update for the U.S. Preventive Services Task Force. Evidence Synthesis No. 58, Part 2. AHRQ Publication No. 08-05108-EF-2. Agency for Healthcare Research and Quality.

Lavrakas, P. J. (2008). Encyclopedia of survey research methods. Sage Publications, Inc.

Lau, J. T., Tsui, H. Y., & Wang, Q. S. (2003). Effects of two telephone survey methods on the level of reported risk behaviours. *Sexually Transmitted Infections*, 79, 325-331.

Ledgerwood, D. M., Goldberger, B. A., Risk, N. K., Lewis, C. E., & Price, R. K. (2008). Comparison between self- report and hair analysis of illicit drug use in a community sample of middle-aged men. *Addicted Behaviors*, 33, 1131-1139.

Lemmens, P., Tan, E. S., Knibbe, R. A. (1992). Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. *Journal of Studies on Alcohol and Drugs*, 53(5), 476-86.

Lessler, J. T., Caspar, R. A., Penne, M. A., & Barker, P. R. (2000). Developing computer assisted interviewing (CAI) for the National Household Survey on Drug Abuse. *Journal of Drug Issues*, 30(1), 9-34.

Levy, S., Sherritt, L., Harris, S. K., Gates, E. C., Holder, D. W., Kulig, J. W., & Knight, J. R. (2004). Test-retest reliability of adolescents' self-report of substance use. *Alcoholism: Experimental and Clinical Research*, 28, 1236-1241.

Lewis, C. C., Matheson, D. H., & Brimacombe, C. A. (2011). Factors influencing patient disclosure to physicians in birth control clinics: an application of the communication privacy management theory. *Health Communication*, 26, 502-511.

Lipari, R. N. & Park-Lee, E. 2019. Key substance use and mental health indicators in the United States: results from the 2018 National Survey on Drug Use and Health. *Substance Abuse and Mental Health Services Administration*.

MacMillan, H. L., Wathen, C. N., Jamieson, E., Boyle, M. H., Shannon, H. S., Ford-Gilboe, M., ... McNutt, L. A., (2009). Screening for Intimate Partner Violence in Health Care Settings: A Randomized Trial. *Journal of the American Medical Association*, 302(5), 493-501.

Madras, B. K., Compton, W. M., Avula, D., Stegbauer, T., Stein, J. B., & Clark, H. W. (2009). Screening, brief interventions, referral to treatment (SBIRT) for illicit drug and alcohol use at multiple healthcare sites: comparison at intake and 6 months later. *Drug and Alcohol Dependence*, 99(1), 280-295.

Magura, S. & Kang, S. Y. (1996). Validity of self-reported drug use in high risk populations: A meta-analytical review. *Substance Use and Misuse*, 31(9), 1131-1153.

Maisto, S., Conigliaro, J. C., Gordon, A. J., McGinnis, K. A., & Justice, A. C. (2008). An experimental study of the agreement of self-administration and telephone administration of the Timeline Followback interview. *Journal of Studies on Alcohol and Drugs*, 69(3), 468-471.

Martin-Willett, R., Helmuth, T., Abraha, M., Bryan, A. D., Hitchcock, L., Lee, K., & Bidwell, L. C. (2020). Validation of a multisubstance online Timeline Followback assessment. *Brain and Behavior*, 10(1), e01486. https://doi.org/10.1002/brb3.1486

Marsch, L. A., Campbell, A., Campbell, C., Chen, C. H., Ertin, E., Ghitza, U., Lambert-Harris, C,. ... & Young, S. (2020). The application of digital health to the assessment and treatment of substance use disorders: The past, current, and future role of the National Drug Abuse Treatment Clinical Trials Network. *Journal of Substance Abuse Treatment,*112S, 4-11.doi: 10.1016/j.jsat.2020.02.005.

McCambridge, J., & Day, M. (2008). Randomized controlled trial of the effects of completing the Alcohol Use Disorders Identification Test questionnaire on self-reported hazardous drinking. *Addiction,* 103(2):241-248.

McCambridge J. & Rollnick, S. (2014). Should brief interventions in primary care address alcohol problems more strongly? *Addiction,* 109, 1054-1058.

McCambridge J. & Saitz, R. (2017). Rethinking brief interventions for alcohol in general Practice. *British Medical Journal,* 356:j116. doi: 10.1136/bmj.j116. PMID: 28108452.

McCann, B. S., Simpson, T. L., Ries, R., & Roy-Byrne, P. (2000). Reliability and validity of screening instruments for drug and alcohol abuse in adults seeking evaluation for attention-deficit/hyperactivity disorder. *American Journal of Addiction,* 9, 1-9.

McLellan, A. T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., & Argeriou, M. (1992). The fifth edition of the addiction severity index. *Journal of Substance Abuse Treatment,* 9, 199-213.

McCual, M. E. & Wand, G. S. (2018). Detecting deception in our research participants: Are your participants who you think they are? *Alcoholism: Clinical and Experiment Research,* 42(2), 230-237.

McKenna, H., Treanor, C., O'Reilly, D., & Donnelly, M. (2018). Evaluation of the psychometric properties of self-reported measures of alcohol consumption: a COSMIN systematic review. *Substance Abuse Treatment, Prevention, and Policy,* 13:6.

McNeely, J., Halkitis, P. N., Horton, A., Khan, R., & Gourevitch, M. N. (2014b). How patients understand the term "nonmedical use" of prescription drugs: insights from cognitive interviews. *Substance abuse*, *35*(1), 12-20.

McNeely, J., Strauss, S. M., Rotrosen, J., Ramautar, A., & Gourevitch, M. N. (2016a). Validation of an audio computer-assisted self-interview (ACASI) version of the alcohol, smoking and substance involvement screening test (ASSIST) in primary care patients. *Addiction,* 111(2), 233-244.

McNeely, J., Strauss, S. M., Saitz, R., Cleland, C .M., Palamar, J. J., & Gourevitch, M. N. (2015a). A brief patient self-administered substance use screening tool for primary care: Two-site validation study of the Substance Use Brief Screen (SUBS). *The American Journal of Medicine*, 128(7), e9-19.

McNeely, J., Cleland, C. M., Strauss, S. M., Palamar, J. J., Rotrosen, J., & Saits, R. (2015b). Validation of Self-Administered Single-Item Screening Questions (SISQs) for Unhealthy Alcohol and Drug Use in Primary Care Patients. *Journal of General Internal Medicine*, 30, 1757-1764.

McNeely, J., Strauss, S. M., Wright, S., Rotrosen, J., Khan, R., Lee, J. D., & Gourevitch, M. N. (2014a). Test–retest reliability of a self-administered Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) in primary care patients. *Journal of Substance Abuse Treatment*, 47(1), 93-101.

McNeely, J., Wu, L. T., Subramaniam, G., Sharma, G., Cathers, L. A., Svikis. D., ... Schwartz, P. (2016b). Performance of the Tobacco, Alcohol, Prescription Medication, and Other Substance Use (TAPS) Tool for Substance Use Screening in Primary Care Patients. *Annals of Internal Medicine*, 165, 690-699.

McNeely, J., Kumar, P. C., Rieckmann, T., Sedlander, E., Farkas, S., Chollak, C., Kannry, J. L., ... Rotrosen, J. (2018). Barriers and facilitators affecting the implementation of substance use screening in primary care clinics: a qualitative study of patients, providers, and staff. *Addiction Science & Clinical Practice*, 13:8.

McPherson, T. L., & Hersch, R. K. (2000). Brief substance use screening instruments for primary care settings: a review. *Journal of Substance Abuse Treatment*, 18(2), 193-202.

McQuade, W. H., Levy, S. M., Yanek, L. R., Davis, S. W., & Liepman, M. R. (2000). Detecting symptoms of alcohol abuse in primary care settings. *Archives of Family Medicine*, 9, 814-821.

Melnikov, A., Hedden, S., & Latimer, W. (2009). Validity of marijuana and opiate use self-report among adult drug users in Novosibirsk, Russia. *Journal of Substance Use*, 14(3-4), 221-229.

Metrik, J., Gunn, R.L., Jackson, K. M., Sokolovsky, A.W., & Borsari, B. (2018). Daily patterns of marijuana and alcohol co-use among individuals with alcohol and cannabis use disorders. *Alcoholism: Clinical and Experimental Research*, 42(6), 1096-1104.

Midanik, L. T. (1994). Comparing usual quantity/frequency and graduated frequency scales to assess yearly alcohol consumption: Results from the 1990 United States National Alcohol Survey. *Addiction*, 89, 407-412.

Mills, T. J., Avegno, J., & Haydel, M. J. (2006). Male victims of partner violence: prevalence and accuracy of screening tools. *Journal of Emergency Medicine*,19(2), 117-24.

Mojtabai, R. & Crum, R. M. (2013). Perceived Unmet Need for Alcohol and Drug Use Treatments and Future Use of Services: Results from a Longitudinal Study. *Drug and Alcohol Dependence*, 127(1-3), 59-64.

Moeller K. E., Lee K. C., & Kissack J. C. (2008). Urine drug screening: practical guide for clinicians. *Mayo Clinic Proceedings*, 83, 66-76.

Moeller, K. E., Kissack, J. C., Atayee, R. S., & Lee, K. C. (2017). Clinical interpretation of urine drug tests: What clinicians need to know about urine drug screens. *Mayo Clinic Proceedings*, 92(5), 774-796.

Moore, R. S. & Ames, G. M. (2002). Survey confidentiality vs. anonymity: Young men's self-reported substance use. *Journal of Alcohol and Drug Education*, 47(2), 32-41.

Moos, R. H. (2008). Context and mechanisms of reactivity to assessment and treatment. *Addiction*, 103(2), 249- 250.

Moyer, V. A. (2013). Screening and behavioral counseling interventions in primary care to reduce unhealthy alcohol use: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine,* 159, 210-218.

Moyer, A., Finney, J. W., Swearingen, C. E. and Vergun, P. (2002). Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. *Addiction*, 97, 279-292.

Mumm, J. & Mutlu, B. (2011). Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. *Computers in Human Behavior*, 27(5), 1643-1650.

Meyers, L. S., Gamst, G. C., & Guarino, A. J. (2017). Applied Multivariate Research: Design and Interpretation – Third Edition. SAGE Publications.

Napper, L. E., Fisher, D. G., Johns, M. E., & Wood, M. M. (2010). The reliability and validity of drug users' self reports of amphetamine use among primarily heroin and cocaine users. *Addicted Behaviors*, 35(4), 350-354.

National Institute on Alcohol Abuse and Alcoholism (NIAAA). (2017). Alcohol facts and statistics.

National Institute on Alcohol Abuse and Alcoholism (NIAAA). (2018). A reduction in the World Health Organization (WHO) Risk levels of alcohol consumption as an efficacy outcome in Alcohol Use Disorder (AUD) Clinical Trials. U.S. Food & Drug Administration Critical Path Innovation Meeting.

National Institute on Drug Abuse (2017). Trends and statistics. Retrieved from https://www.drugabuse.gov/related-topics/trends-statistics

National Survey on Drug Use and Health (2015). America's need for and receipt of substance use treatment in 2015. Retrieved from https://www.samhsa.gov/data/sites/default/files/report_2716/ShortReport-2716.html

Neal, J. & Robertson, M. (2003). Comparisons of self-report data and oral fluid testing in detecting drug use amongst new treatment clients. *Drug and Alcohol Dependence*, 71(1), 57-64.

Newcombe, D. A. L., Humeniuk, R. E., & Ali, R. (2005). Validation of the world health organization alcohol, smoking and substance involvement screening test (ASSIST). Report of results from the Australian site. *Drug and Alcohol Review*, 24, 217-26.

Newman, J. C., Des Jarlais, D. C., Turner, C. F., Gribble, J., Cooley, P., & Paone, D. (2002). The differential effects of face-to-face and computer interview modes. *American Journal of Public Health*, 92, 294-297.

National Institute on Drug Abuse (NIDA). (2020). Sex and Gender Differences in Substance Use. Retrieved from https://www.drugabuse.gov/publications/research-reports/substance-use-in-women/sex-gender-differences-in-substance-use.

NIMH Multisite HIV/STD Prevention Trial for African American Couples Group. (2008). Designing an audio computer-assisted self-interview (ACASI) system in a multi-site trial: A brief report. *Journal of Acquired Immune Deficiency Syndromes*, 49, S52-S58.

Norberg, M. M., Mackenzie, J., & Copeland, J. (2012). Quantifying cannabis use with the timeline followback approach: a psychometric evaluation. *Drug and Alcohol Dependence*, 121(3), 247-52.

Nordeck, C. D., Gryczynski, J., O'Grady, K.E., Polak, K., Svikis, D. S., McNeely, J., Wu, L., & Schwartz, R. P. (2020). Comparison of timeline follow-back self-report and oral fluid testing to detect substance use in adult primary care patients. *Drug and Alcohol Dependence*, 209(1), 107939.

Ochalek, T. A., Cumpston, K. L., Wills, B. K., Gal, T. S., & Moeller, G. (2020). Nonfatal opioid overdoses at an urban emergency department during the Covid-19 pandemic. *Journal of the American Medical Association*, e2017477. doi: 10.1001/jama.2020.17477.

O'Connor, E. A., Perdue, L. A., Senger, C. A., Ruskin, M., Patnode, C. D., Bean, S. I., & Jonas, D. E. (2018). Screening and Behavioral Counseling Interventions to Reduce Unhealthy Alcohol Use in Adolescents and Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *Journal of the American Medical Association*, 320(18),1910-1928.

O'Hare, T., Bennett, P., & Leduc, D. (1991). Reliability of self–reports of alcohol use by community clients. *Hospital & Community Psychiatry*, 42, 406-408.

Olsen, Y. & Sharfstein, J. M. (2014). Confronting the stigma of opioid use disorder--and its treatment. *Journal of the American Medical Association*, 311(14), 1393-1394.

O'Malley, P. M., Johnston, L. D., Bachman, J. G., & Schulenberg, J. (2000). A comparison of confidential versus anonymous survey procedure: Effects on reporting of drug use and related attitudes and beliefs in a national study of students. *Journal of Drug Issues*, 30(1), 35-54.

Ondersma, S. J., Svikis, D. S., Lebreton, J. M., Streiner, D. L., Grekin, E. R, Lam, P. K, & Connors-Burge, V. (2012). Development and preliminary validation of an indirect screener for drug use in the perinatal period. *Addiction*, 107(12), 2099-2106.

Ong, A. D. & Weiss, D. J. (2000). The impact of anonymity on responses to sensitive questions. *Journal of Applied Social Psychology*, 30, 1691-1708.

Pace, C. A. & Uebelacker, L. A. (2018). Addressing unhealthy substance use in primary care. *Medical Clinics of North America*, 102(4), 567-586.

Park-Lee, E., Lipari, R. N., Hedden, S. L., Kroutil, L. A., & Porter, J. D. (2017). Receipt of services for substance use and mental health issues among adults: Results from the 2016 National Survey on Drug Use and Health. *NSDUH Data Review*.

Patnode, C. D., Perdue, L. A., Rushkin, M., & O'Connor, E. A. (2019). Screening for Drug Use in Primary Care in Adolescents and Adults, Including Pregnant Women: An Updated Systematic Review for the U.S. Preventive Services Task Force: Evidence Synthesis No. 186. Agency for Healthcare Research and Quality. AHRQ publication 19-05255-EF-1.

Patnode, C. D., Perdue, L. A., Rushkin, M., Dana, T., Blazina, I, Bougatsos, C., Grusing, S., ... & Chou, R. (2020). Screening for Unhealthy Drug Use: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *Journal of the American Medical Assocation*,323(22), 2310-2328.

Peacock, A., Leung, Larney, S., Colledge, S., Hickman, M., Rehm, ... Degenhardt, L. (2018). Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. *Addiction*, 113, 1905-1926.

Pedersen, E. R., Grow, J., Duncan, S., Neighbors, C., & Larimer, M. E. (2012). Concurrent validity of an online version of the timeline Followback assessment. *Psychology of Addictive Behaviors*, 26(3), 672-677.

Perlis, T. E., Des Jarlais, D. C., Friedman, S. R., Arasteh, K., & Turner, C. F. (2004). Audio-computerized self-interviewing versus face-to-face interviewing for research data collection at drug abuse treatment programs. *Addiction*, 99(7), 885-896.

Perron, B. E., Mowbray, O. P., Glass, J. E., Delva, J., Vaughn, M. G., & Howard, M. O. (2009). Differences in service utilization and barriers among blacks, Hispanics, and whites with drug use disorders. *Substance Abuse Treat and Prevention Policy*, 13, 4:3.

Pierce, B. S., Perrin, P. B., Tyler, C. M., McKee, G. B., & Watson, J. D. (2020). The COVID-19 telepsychology revolution: A national study of pandemic-based changes in U.S. mental health care delivery. *American Psychologist.* Advance online publication. http://dx.doi.org/10.1037/amp0000722

Pilowsky, D. J., & Wu, L. T. (2012). Screening for alcohol and drug use disorders among adults in primary care: a review. *Substance abuse and rehabilitation*, *3*(1), 25-34.

Preston, K. L., Silverman, K., Schuster, C. R., & Cone, E. J. (1997). Comparison of self-reported drug use with quantitative and qualitative urinalysis of drug use in treatment studies. Monograph 167, National Institute on Drug Abuse, 130-145.

Pokorny, A. D., Miller, B. A., & Kaplan, H. B. (1972). The brief MAST: A shortened version of the Michigan Alcoholism Screening Test. *The American Journal of Psychiatry*, 129(3), 342-345.

Polak, K., Haug, N. A., Drachenberg, H. E., & Svikis, D. S. (2015). Gender considerations in addiction: implications for treatment. *Current Treatment Options in Psychiatry*, 2(3), 326-338.

Polak, K., Kelpin, S., Reisweber, J., & Svikis, D. S. (2020). Evidence-Based Behavioral Therapies for Substance Use Disorders. In Moeller, G. & Terplan, M. (Eds). Primer on Substance Use Disorders (pp. 267-281). Oxford University Press. doi: 10.1093/med/9780190920197.001.0001

Pollard, M. S., Tucker, J. S., & Green, H. D. (2020). Changes in Adult Alcohol Use and Consequences During the COVID-19 Pandemic in the US. *JAMA Network Open*, 3(9), e2022942. doi:10.1001/jamanetworkopen.2020.22942

Pomerleau, C. S., Carton, S. M., Lutzke, M. L., Flessland, K. A., & Pomerleau, O. F. (1994). Reliability of the Fagerström tolerance questionnaire and the Fagerström test for nicotine dependence. *Addictive Behaviors*, 19(1), 33-39.

Raistrick, D., Braddshaw, J., Tober, G., Weiner, J., Allison, J., & Healey, C. (1994). Development of the Leeds Dependence Questionnaire (LDQ): A questionnaire to measure alcohol and opiate dependence in the context of a treatment evaluation package. *Addiction*, 89, 563-572.

Rehm, J. (2020). Alcohol use is a key factor in recent decreases in life expectancy in the United States. *Alcoholism: Clinical and Experimental Research*, 44(2), 404-406.

Rehm, J., Anderson, P., Manthey, J., Shield, K. D., Struzzo, P., Wojnar, M., Gual, A. (2016). Alcohol use disorders in primary health care: what do we know and where do we go? *Alcohol and Alcoholism*, 51(4), 422-7.

Rehm, J., Greenfield, T. K., & Rogers, J. D. (2001). Average volume of alcohol consumption, patterns of drinking, and all-cause mortality: Results from the US National Alcohol Survey. *American Journal of Epidemiology*, 153(1), 64-71.

Rendon, A., Livingston, M., Suzuki, S., Hill, W., & Walters, S. (2017). What's the agreement between self-reported and biochemical verification of drug use? A look at permanent supportive housing residents. *Addicted Behaviors*, 70, 90-96.

Reus, V. I, Fochtmann, L. J., Bukstein, O., Eyler, A. E., Hilty, D. M., Horvitz-Lennon, M., ... Hong, S. H. (2018). The American Psychiatric Association practice guideline for the pharmacological treatment of patients with alcohol use disorder. *American Journal of Psychiatry*, 175, 86-90.

Richman, W. L., Weisband, S., Kiesler, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, 84(5), 754-775.

Richter, L. & Johnson, P. B. (2001). Current methods of assessing substance use: A review of strengths, problems, and developments. *Journal of Drug Issues*, 31, 809-832.

Riley, K. J., Lu, N. T., & Taylor, B. G. (2000). Drug screening: A comparison of urinalysis results from two independent laboratories. *Journal of Drug Issues*, 30(1), 171-185.

Robins, L. N., Wing, J., Wittchen, H. U., Helzer, J. E., Babor, T. F., Burke, J., ... Towle, L.H. (1988). The Composite International Diagnostic Interview: An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry*, 45, 1069-1077.

Robinson, S. M., Sobell, L. C., Sobell, M. B., & Leo, G. I. (2014). Reliability of the Timeline Followback for cocaine, cannabis, and cigarette use. *Psychology of Addictive Behaviors*, 28, 154-162.

Rogelberg, S. C., Spitzmüeller, C., Little, I., & Reeve, C. L. (2006). Understanding response behavior to an online special topics organizational satisfaction survey. *Personnel Psychology*, 59, 903-923.

Rohn, M. C., Lee, M. R., Kleuter, S. B., Schwandt, M. L., Falk, D. E, & Leggio, L. (2017). Differences between treatment-seeking and nontreatment-seeking alcohol-dependent research participants: An exploratory analysis. *Alcoholism: Clinical and Experimental Research*, 41(2), 414-420.

Room, R. (1990) Measuring alcohol consumption in the United States: methods and rationales. In *Research Advances in Alcohol and Drug Problems*, Vol. 10, Kozlowski, L. T., Annis, H. M., Cappel, H. D., Glaser, F. B., Goodstadt, M. S., Israel, Y., Kalant, H., ... & Vingilis, E. R. (Eds.), pp. 39–80. Plenum Press.

Rosenbaum, A., Rabenhorst, M. M., Reddy, M. K., Fleming, M. T., & Howells, N. L. (2006). A comparison of methods for collecting self-report data on sensitive topics. *Violence and Victims*, 21, 461-471.

Roy, M., Dum, M., Sobell, L. C., Sobell, M. B., Simco, E. R., Manor, H., & Palmerio, R. (2008). Comparison of the quick drinking screen and the alcohol timeline followback with outpatient alcohol abusers. *Substance Use & Misuse*, 43(14), 2116-2123.

Roy-Byrne, P., Bumgardner, K., Krupski, A., Dunn, C., Ries, R., Donovan, D., ... Zarkin, G A. (2014). Brief intervention for problem drug use in safety-net primary care settings: a randomized clinical trial. *Journal of the American Medical Association*, 312, 492-501.

Rudd, R. A., Aleshire, N., Zibbel, J. E., & Gladden, R. M. (2016). Increases in drug and opioid overdose deaths — United States, 2000–2014. *Morbidity and Mortality Weekly Report (MMWR)*, 64(50), 1378-1382.

Rueger, S. Y., Trela, C. J., Palmeri, M., & King, A. C. (2012). Self-administered web-based Timeline Followback procedure for drinking and smoking behaviors in young adults. *Journal of Studies on Alcohol and Drugs*, 73(5), 829-833.

Russell, M., Light, J. M., & Gruenewald, P. J. (2004). Alcohol consumption and problems: The relevance of drinking patterns. *Alcoholism: Clinical and Experimental Research*, 28(6), 921-930.

Sacks, J. J., Gonzales, K. R., Bouchery, E. E., Tomedi, L. E., & Brewer, R. D. (2015). 2010 national and state costs of excessive alcohol consumption. *American Journal of Preventive Medicine*, 49(5), e73-e79.

Sacks, J. A., Drake, R. E., Williams, V. F., Banks, S. M., & Herrell, J. M. (2003). Utility of the Time-Line Follow-Back to assess substance use among homeless adults. *Journal of Nervous and Mental Disease*, 191, 145-153.

Sahker, E., Jones, D., Lancianese, D. A., Pro, G., & Arndt, S. (2019). Racial/Ethnic differences in alcohol and drug use outcomes following Screening, Brief Intervention, and Referral to Treatment (SBIRT) in Federally Qualified Health Centers. *Journal of Racial and Ethnic Health Disparities*, 6(6), 1192-1199.

Saitz, R. (2007). Screening and brief intervention enter their 5th decade. *Substance Abuse*, 28(3), 3-6.

Saitz, R. (2010). Alcohol screening and brief intervention in primary care: Absence of evidence for efficacy in people with dependence or very heavy drinking. *Drug and Alcohol Review*, *29*(6), 631-640.

Saitz, R. (2013). Addressing Unhealthy Alcohol Use in Primary Care. Springer.

Saitz R. (2014). Screening and brief intervention for unhealthy drug use: Little or no efficacy. *Frontiers in Psychiatry*, 5:121.

Saitz, R., & Daaleman, T. P. (2017). Now is the time to address substance use disorders in primary care. *Annals of Family Medicine*, 15(4), 306-308.

Saitz, R., Cheng, D. M., Allensworth-Davies, D., Winter, M.R., & Smith, P. C. (2014b). The ability of single screening questions for unhealthy alcohol and other drug use to identify substance dependence in primary care. *Journal of Studies on Alcohol and Drugs*, 75, 153-157.

Saitz, R., Palfai, T. P., Cheng, D. M., Alford, D. P., Bernstein, J. A., Lloyd-Travaglini, C. A., ... & Samet, J. H. (2014a). Screening and brief intervention for drug use in primary care: the ASPIRE randomized clinical trial. *Journal of the American Medical Association*, *312*(5), 502-513.

Samet, S., Waxman, R., Hatzenbuehler, M., & Hasin, D. S. (2007). Assessing addiction: Concepts and instruments. *Addiction Science & Clinical Practice*, 4(1), 19-31.

Sankar, P., Mora, S., Merz, J.F., & Jones, N.L. (2003). Patient perspectives of medical confidentiality: A review of the literature. *Journal of General Internal Medicine*, 18(8), 659-669.

Sarkar, M., Einarson, T., & Koren, G. (2010) Comparing the effectiveness of TWEAK and T-ACE in determining problem drinkers in pregnancy. *Obstetrics & Gynecology*, 116(4), 827-33.

Saunders, J. B. & Conigrave, K. M. (1990). Early identification of alcohol problems. *Canadian Medical Association Journal*, 143, 1060-1069.

Saunders, J. B., Aasland, O. G., Babor, T. E, De La Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on early detection of persons with harmful alcohol consumption. *Addiction*, 88, 791-804.

Sayre, M., Lapham, G. T., Lee, A. K., Oliver, M., Bobb, J. F., Caldeiro, R. M., & Bradley, K. A. (2020). Routine Assessment of Symptoms of Substance Use Disorders in Primary Care: Prevalence and Severity of Reported Symptoms. *Journal of General Internal Medicine*, 35, 1111–1119.

Scholl, L., Seth, P., Kariisa, M., Wilson, N., & Baldwin G. (2019). Drug and Opioid-Involved Overdose Deaths – United States, 2013-2017. *Morbidity and Mortality Weekly Report (MMWR)*, 67, 1419-1427.

Schomerus, G., Lucht, M., Holzinger, A., Matschinger, H., Carta, M. G., & Angermeyer, M. C. (2011). The stigma of alcohol dependence compared with other mental disorders: A review of population studies. *Alcohol and Alcoholism*, 46(2), 105-112.

Schuler, M., Lechner, W., Carter, R., & Malcolm, R. (2009). Temporal and gender trends in concordance of urine drug screens and self-reported use in cocaine treatment studies. *Journal of Addiction Medicine*, 3(4), 211-217.

Scott, C. R. (2005). Anonymity in applied communication research: Tension between IRBs, researchers, and human subjects. *Journal of Applied Communication Research*, 33, 242-257.

Scott, K. M., Lim, C., Al-Hamzawi, A., Alonso, J., Bruffaerts, R., Caldas-de-Almeida, J. M., & de Jonge, P. (2016). Association of mental disorders with subsequent chronic physical conditions: World mental health surveys from 17 countries. *JAMA Psychiatry*, 73(2), 150-158.

Seale, J. P., Boltri, J. M., Shellenberger, S., Velasquez, M. M., Cornelius, M., Guyinn, M., Okosun, I., & Sumner, H. (2006). Primary care validation of a single screening question for drinkers. *Journal of Studies on Alcohol*, 67(5), 778-784.

Searles, J. S., Helzer, J. E., & Walter, D. E. (2000). Comparison of drinking patterns mea sured by daily reports and timeline follow back. *Psychology of Addictive Behaviors*, 14(3), 277-286.

Secades-Villa, R. & Fernandez-Hermida, J. R. (2003). The validity of self-reports in a follow-up study with drug addicts. *Addicted Behaviors*, 28(6), 1175-1182.

Sellers, E. M., Somer, G. R., Sobell, L., & Sobell, M. (1990). Discontinuation symptoms among persistent alprazolam users. *Clinical Pharmacology and Therapeutics*, 47, 168.

Selzer, M. L. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument. *American Journal of Psychiatry*, 127, 1653-1658.

Selzer, M. L. Vinokur, A., & Van Rooijen, L. (1975). A self-administered Short Michigan Alcoholism Screening Test. *Journal of Studies on Alcohol*, 36(1), 117-126.

Shapiro, B., Coffa, D., & McCance-Katz, E. F. (2013). A primary care approach to substance misuse. *American Family Physician*, 88(2) (2013), 113-121.

Sherman, M. F. & Bigelow, G. E. (1992). Validity of patients' self-reported drug use as a function of treatment status. *Drug and Alcohol Dependence*, 30, 1-11.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2):420.

Shupp, R., Loveridge, S., Skidmore, M., Green, B., & Albrecht, D. (2020). Recognition and stigma of prescription drug abuse disorder: Personal and community determinants. *BMC Public Health*. doi: https://doi.org/10.21203/rs.3.rs-15664/v2.

Simões, A. M. & Bastos, F. I. (2004). Audio Computer-Assisted Interview: a new technology in the assessment of sexually transmitted diseases, HIV, and drug use. *Cadernos de Saúde Pública*, 20, 1169-1181.

Simões, A. A., Bastos, F. I., Moreira, R. I., Lynch, K. G., & Metzger, D. S. (2006). A randomized trial of audio computer and in-person interview to assess HIV risk among drug and alcohol users in Rio De Janeiro, Brazil. *Journal of substance abuse treatment*, 30(3), 237-243.

Skinner, H. A. (1982). The Drug Abuse Screening Test. *Addictive Behaviors*, 7(4), 363-367.

Skinner, H. G., Coffey, R., Jones, J., Heslin, K. C., & Moy, E. (2016). The effects of multiple chronic conditions on hospitalization costs and utilization for ambulatory care sensitive conditions in the United States: a nationally representative cross-sectional study. *BMC Health Service Research*, 16, 77. doi: https://doi.org/10.1186/s12913-016-1304-y

Smith, P. C., Schmidt, S. M., Allensworth-Davies, D., and Saitz, R. (2009). Primary care validation of a single-question alcohol screening test. *Journal of General Internal Medicine*, 24, 783-788.

Smith, P. C., Schmidt, S. M., Allensworth-Davies, D., and Saitz, R. (2010). A Single-Question Screening Test for Drug Use in Primary Care. *Archive of Internal Medicine*, 170(13), 1155-1160.

Smith, M. W., Stocks, C., & Santora, P. B. (2015). Hospital readmission rates and emergency department visits for mental health and substance abuse conditions. *Journal of Community Mental Health*, 51, 190-197

Smith, M. T. & Wegener, S. T. (2003). Measures of sleep: The Insomnia Severity Index, Medical Outcomes Study (MOS) Sleep Scale, Pittsburgh Sleep Diary (PSD), and Pittsburgh Sleep Quality Index (PSQI). *Arthritis & Rheumatology*, 49(Suppl 5), S184-96.

Sobell, L. C. & Sobell, M. B. (1992). Timeline follow-back: A technique for assessing self-reported alcohol consumption. In R.Z. Litten & J. Allen (Eds.), *Measuring alcohol consumption*: *Psychosocial and biological methods* (p. 41-72*).* Humana Press.

Sobell, L. C. & Sobell, M. B. (2008). Alcohol Timeline Followback (TFLB). In American Psychiatric Association (Ed.), Handbook of Psychiatric Measures (p. 477-479). American Psychiatric Association.

Sobell, L. C. & Sobell, M. B. (2003). Alcohol consumption measures. In: Allen, J.P. and Wilson, V. *Assessing Alcohol Problems, 2nd Edition (*p. 78-79). National Institute on Alcohol Abuse and Alcoholism.

Sobell, L. C., Agrawal, S., Annis, H., Ayala-Velazquez, H., Echeverria, L., Leo, G. I., & Ziólkowski, M. (2001). Cross-cultural evaluation of two drinking assessment instruments: Alcohol timeline followback and inventory of drinking situations. *Substance Use & Misuse*, 36(3), 313-331.

Sobell, L. C., Agrawal, S., Sobell, M. B., Leo, G. I., Young, L. J., Cunningham, J. A., and Simco, E. R. (2003). Comparison of a Quick Drinking Screen with the Timeline Followback for Individuals with Alcohol Problems. *Journal of Studies on Alcohol*, 64(6), 858-861.

Sobell, L. C. & Sobell, M. S. (2004). Alcohol consumption measures. Retrieved at: http://pubs.niaaa.nih.gov/publications/AssessingAlcohol/measures.htm

Sobell, L. C., Brown, J., Leo, G. I., & Sobell, M. B. (1996). The reliability of the Alcohol Timeline Followback when administered by telephone and by computer. *Drug and Alcohol Dependence*, 42, 49-54.

Solberg, L. I., Maciosek, M. V., & Edwards, N. M. (2008). Primary care intervention to reduce unhealthy alcohol use ranking its health impact and cost effectiveness. *American Journal of Preventive Medicine*, 34(2), 143-152.

Sokol, R. J., Martier, S. S., & Ager, J. W. (1989). The T-ACE questions: Practical prenatal detection of risk drinking. *American Journal of Obstetrics and Gynecology*, 160, 863-871.

Spear, S. E., Shedlin, M., Gilberti, B., Fiellin, M., & McNeely, J. (2016). Feasibility and Acceptability of an Audio Computer-Assisted Self-Interview Version of the Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST) in Primary Care Patients. *Substance Abuse*, 37(2), 299–305.

Stahre, M., Roeber, J., Kanny, D., Brewer, R. D., & Zhang, X. (2014). Contribution of excessive alcohol consumption to deaths and years of potential life lost in the United States. *Preventing Chronic Disease*, 11(E109).

Staines, G. L., Magura, S., Foote, J., Deluca, A., & Kosanke, N. (2001). Polysubstance use among alcoholics. *Journal of Addictive Diseases*, 20(4), 53-69.

Staudt, A., Freyer-Adam, J., Meyer, C., John, U., & Baumann, S (2018). Short-term stability of different drinking patterns over the course of four weeks among adults. A latent transition analysis. *Drug and Alcohol Dependence,* 191(1), 181-186.

Stefanidou, M., Athanaselis, S., Spiliopoulou, C., Dona, A., & Maravelias, C. (2010). Biomarkers of opiate use. *Internal Journal of Clinical Practice*, 64, 1712-1718.

Steinbauer, J. R., Cantor, S. B., Holzer, C. E., & Volk, R. J. (1998). Ethnic and sex bias in primary care screening tests for alcohol use disorders. *Annals of Internal Medicine,* 129(5), 353-362.

Steinweg, D. L. & Worth, H. (1993). Alcoholism: The keys to the CAGE. *American Journal of Medicine*, 94, 520-523.

Stockwell, T., Donath, S., Cooper-Stanbury, M., Chikritzhs, T.N., Catalano, P., & Mateo, C. (2004). Under-reporting of alcohol consumption in household surveys: a comparison of quantity-frequency, graduate-frequency and recent recall. *Addiction,* 99(8), 1024-33.

Strack, F. (1992). "Order effects" in survey research: Activation and information functions of preceding questions. In Schwartz, N. & Sudman, S. (Eds.), *Context effect in social and psychological research* (pp. 23-34). *Springer-Verlag.*

Strobbe, S. (2014). Prevention and screening, brief intervention, and referral to treatment for substance use in primary care. *Primary Care*, 41, 185-213.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2006). National Survey on Drug Use and Health. (Office of Applied Studies, NSDUH Series H-32, DHHS Publication No. SMA 07-4293).

Substance Abuse and Mental Health Services Administration (SAMHSA). (2010). Reliability of Key Measures in the National Survey on Drug Use and Health (Office of Applied Studies, Methodology Series M-8, HHS Publication No. SMA 09-4425).

Substance Abuse and Mental Health Services Administration (SAMHSA). (2012). The Role of Biomarkers in the Treatment of Alcohol Use Disorders, Revision. Advisory Councils, Volume 11, Issue 2.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2019). Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health (HHS Publication No. PEP19-5068, NSDUH Series H-54). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2020a). Screening, Brief intervention, and Referral to Treatment (SBIRT). Office of National Drug Control Policy ONDCP.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2020b). The Opioid Crisis and the Black/African American Population: An Urgent Issue. Publication No. PEP20-05-02-001. Office of Behavioral Health Equity. Substance Abuse and Mental Health Services Administration.

Svikis, D. S., McCaul, M. E., Turkkan, J. S. & Bigelow, G. E. (1991). Effects of item correction on Michigan Alcoholism Screening Test scores in college men with and without a family history of alcoholism. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*(4), 654-659.

Svikis, D. S., Keyser-Marcus, L., Stitzer, M., Rieckmann, T., Safford, L., Loeb, P., … Zweben J. (2012). Randomized multi-site trial of the Job Seekers' Workshop in patients with substance use disorders. *Drug and Alcohol Dependence,* 120, 55-64.

Svikis, D. S. & Reid-Quiñones, K. (2003). Screening and prevention of alcohol and drug use disorders in women. *Obstetrics and Gynecology Clinics of North America, 30*(3), 447-68.

Tai, B., Wu, L. T., & Clark, H. W. (2012). Electronic health records: essential tools in integrating substance abuse treatment with primary care. *Substance Abuse and Rehabilitation,* 3, 1-8.

Tan, C. H., Hungerford, D. W., Denny, C. H., & McKnight-Eily, L. R. (2016). Screening for unhealthy alcohol use: Practices among U.S. primary care providers, DocStyles. *American Journal of Preventive Medicine*, 54(2), 173-180.

Tiet, Q. Q., Finney, J. W., & Moos, R. H. (2008). Screening psychiatric patients for illicit drug use disorders and problems. *Clinical Psychology Review,* 28(4), 578-591.

Tiet, Q. Q., Leyva, Y. E., Moos, R. H., Frayne, S. M, Osterberg, L., & Smith, B. (2015). Screen of Drug Use: Diagnostic Accuracy of a New Brief Tool for Primary Care. *JAMA Intern Medicine*, 175(8), 1371-1377.

Trevethan, R. (2017). Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Frontiers in Public Health Epidemiology*, 5, 307. doi: 10.3389/fpubh.2017.00307

Toll, B. A., Leeman, R. F., McKee, S. A., & O'Malley, S. S. (2008). A brief 7-day estimate of alcohol consumption for use in smoking cessation clinical trials. *Addictive Behaviors*, 33(12), 1606-1609.

Tong, S. T., Polak, K. M., Weaver, M. F., Villalobos, G. C., Smith, W. R., & Svikis, D. S. (2019). Screening for Psychotherapeutic Medication Misuse in Primary Care Patients: Comparing Two Instruments. *Journal of the American Board of Family Medicine,* 32(2), 272–278.

Tonigan, J. S. & Miller, W. R. (2002). The Inventory of Drug Use Consequences (InDUC): Test-retest stability and sensitivity to detect change. *Psychology of Addictive Behaviors*, 16(2), 165-168.

Tourangeau, R. & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859-883.

Tourangeau, R., Rasinski, K., Jobe, J. B., Smith, T. W., & Pratt, W. F. (1997). Sources of error in a survey on sexual behavior. *Journal of Official Statistics*, 13(4), 341-365.

U.S. Department of Health and Human Services (HHS). (2016). Office of the Surgeon General, Facing Addiction in America: The Surgeon General's Report on Alcohol, Drugs, and Health. HHS.

U.S. Preventive Services Task Force (USPSTF). (2018). *Unhealthy Alcohol Use in Adolescents and Adults: Screening and Behavioral Counseling Interventions.* Agency for Healthcare Research and Quality.

U.S. Preventive Services Task Force (USPSTF). (2019). *Draft Recommendation Statement Illicit Drug Use, Including Nonmedical Use of Prescription Drugs: Screening.* Agency for Healthcare Research and Quality.

U.S. Preventive Services Task Force (USPSTF). (2020). Screening for Unhealthy Drug Use: U.S. Preventive Services Task Force Recommendation Statement. *Journal of the American Medical Association*, 323(22), 2301-2309.

Ustun, B., Compton, W., Mager, D., Babor, T., Baiyewu, O., Chatterji, S., ... Sartorius, N. (1997). WHO study on the reliability and validity of the alcohol and drug use disorder instruments. *Drug and Alcohol Dependence*, 47, 161-169.

Vakili, S., Sobell, L. C., Sobell, M. B., Simco, E. R., & Agrawal, S. (2008). Using the Timeline Followback to determine time windows representative of annual alcohol consumption with problem drinkers. *Addictive Behaviors*, 33(9), 1123-1130.

van Boekel, L. C., Brouwers, E. P., van Weeghel, J., & Garretsen, H. F. (2013). Stigma among health professionals towards patients with substance use disorders and its consequences for healthcare delivery: systematic review. *Drug and Alcohol Dependence*, 131(1-2), 23-35.

van Griensven, F., Naorat, S., Kilmarx, P. H., Jeeyapant, S., Manopaiboon, C., Chaikummao, S., ... Tappero, J. W. (2006). Palmtop-assisted self-interviewing for the collection of sensitive behavioral data: Randomized trial with drug use urine testing. *American Journal of Epidemiology*, 163, 271-278.

Viel, G., Boscolo-Berto, R., Cecchetto, G., Fais, P., Nalesso, A., & Ferrara, S. D. (2012). Phosphatidylethanol in Blood as a Marker of Chronic Alcohol Use: A Systematic Review and Meta-Analysis. *International Journal of Molecular Sciences*, 13(11), 14788-14812.

Vigil-Colet, A., Lorenzo-Seva, U., & Morales-Vives, F. (2015). The effects of ageing on self-reported aggression measures are partly explained by response bias. *Psicothema*, 27, 209-215.

Walter, K. N., Wagner, J. A., Cengiz, E., Tamborlane, W. V., & Petry, N. M. (2017). Substance use disorders among patients with type 2 diabetes: a dangerous but understudied combination. *Current Diabetes Reports*, 17(1):2. doi: 10.1007/s11892-017-0832-0. PMID: 28101793.

Wayal S., Reid, D., Blomquist, P. B., Weatherburn, P., Mercer, C. H., & Hughes, G. (2018). The acceptability and feasibility of implementing a bio-behavioral enhanced surveillance tool for sexually transmitted infections in England: Mixed-Methods study. *JMIR Public Health Surveillance*, 4(2):e52 10.2196/publichealth.9010.

Welte, J. W. & Russell, M. (1993). Influence of socially desirable responding in a study of stress and substance abuse. *Alcoholism: Clinical and Experimental Research*, 17, 758-761.

White, B., Day, C. & Maher, L. (2007). Self reported risk behaviour among injecting drug users: Self versus assisted questionnaire completion. *AIDS Care*, 19(3), 441-447.

White, D., Rosenberg, E. S., Cooper, H. L., del Rio, C., Sanchez, T. H., Salazar, L.F., & Sullivan, P. S. (2014). Racial differences in the validity of self-reported drug use among men who have sex with men in Atlanta, GA. *Drug and Alcohol Dependence*, 138, 146-153.

Williams, E. C., Johnson, M. L., Lapham, G. T., Caldeiro, R. M., Chew, L., Fletcher, G. S., ... Bradley, K. A. (2011). Strategies to implement alcohol screening and brief intervention in primary care settings: A structured literature review. *Psychology of Addictive Behaviors*, 28(2), 206-214

Williams, M. L., Freeman, R. C., Bowen, A. M., Zhao, Z., Ellwood, W. N., Gordan, C., Young, P., ... & Sgnes, C-A. (2000). A comparison of the reliability of self-reported drug use and sexual behaviors using computer-assisted versus face-to-face interviewing. *AIDS Education and Prevention*, 12(3), 199-213.

Wilson, N., Kariisa, M., Seth, P., Smith, H., & Davis, N.L. (2020). Drug and opioid-involved overdose deaths – United States, 2017-2019. *Morbidity and Mortality Weekly Report (MMWR)*, 69, 290-297.

Winhusen, T., Theobald, J., Kaelber, D. C., & Lewis, D. (2019). Medical complications associated with substance use disorders in patients with type 2 diabetes and hypertension: Electronic health record findings. *Addiction*, 114(8), 1462-1470.

Wish, E. D., Hoffman, J. A., & Nemes, S. (1997). The validity of self-reports of drug use at treatment admission and at followup: Comparison with urinalysis and hair assays. In L. Harrison, A. Hughes (Eds.), The validity of self-reported drug use: Improving the accuracy of survey estimates, U.S. Department of Health and Human Services, National Institute on Drug Abuse, pp. 200-226.

Wolff, N. & Shi, J. (2015). Screening for Substance Use Disorder Among Incarcerated Men with the Alcohol, Smoking, Substance Involvement Screening Test (ASSIST): A Comparative Analysis of Computer-Administered and Interviewer-Administered Modalities. *Journal of Substance Abuse Treatment*, 53, 22-32.

World Health Organization (WHO) ASSIST Working Group. (2002). The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. *Addiction*, 97, 1183-1194.

Wu, L. T., Gersing, K. R., Swartz, M. S., Burchett, B., Li, T. K., & Blazer, D. G. (2013). Using electronic health records data to assess comorbidities of substance use and psychiatric diagnoses and treatment settings among adults. *Journal of Psychiatric Research*, 47, 555-563.

Wu, L. T., Ghitza, U. E., Zhu, H., Spratt, S., Swartz, M., & Mannelli, P. (2018). Substance use disorders and medical comorbidities among high-need, high-risk patients with diabetes. *Drug and Alcohol Dependence*, 186, 86-93.

Wu, L., McNeely, J., Subramaniam, G. A., Brady, G. T., Sharma, G., Velduisen, P., Zhu, H., Schwartz, R. P. (2017). DSM-5 substance use disorders among adult primary care patients: Results from a multisite study. *Drug and Alcohol Dependence*, 179(1), 42-46.

Yang, L. H., Wong, L. Y., Grivel, M. M., & Hasin, D. S. (2017). Stigma and substance use disorders: an international phenomenon. *Current Opinion in Psychiatry*, 30(5), 378-388.

Young, M. M., Stevens, A., Galipeau, J., Pirie, T., Garritty, C., Singh, K., … Moher, D. (2014). Effectiveness of brief interventions as part of the Screening, Brief Intervention and Referral to Treatment (SBIRT) model for reducing the nonmedical use of psychoactive substances: A systematic review. *Systematic Reviews*, 3(1):1.

Yudko. E., Lozhkina, O., & Fouts, A. (2007). Comprehensive Review of Psychometric Properties of the Drug Abuse Screening Test. *Journal of Substance Abuse Treatment*, 32(2), 189-198.

Zanis, D. A., McLellan, A. T., & Randall, M. (1994). Can you trust patient self-reports of drug use during treatment? *Drug and Alcohol Dependence*, 35, 127-132.

Zarkin, G., Bray, J., Hinde, J., & Saitz, R. 2015. Costs of screening and brief intervention for illicit drug use in rimary Care settings. *Journal of Studies on Alcohol and Drugs*, 76(2), 222-228.

Zung, B. J. (1982). Evaluation of the Michigan Alcoholism Screening Test (MAST) in assessing lifetime and recent problems. *Journal of Clinical Psychology*, 38(2), 425-439.

Appendix A

TIMELINE FOLLOWBACK CALENDAR: 2018

1 Standard Drink is Equal to

One 12 oz can/bottle of beer

One 5 oz glass of regular (12%) wine

1 ½ oz of hard liquor (e.g. rum, vodka, whiskey)

1 mixed or straight drink with 1 ½ oz hard liquor

Complete the Following

Start Date (Day 1): _____ **End Date** (yesterday): _____

MO DY YR MO DY YR

2018	SUN	MON	TUES	WED	THURS	FRI	SAT
		1 New Year's	2	3	4	5	6
J	7	8	9	10	11	12	13
A	14	15 M. Luther King	16	17	18	19	20
N	21	22	23	24	25	26	27
	28	29	30	31	1	2	3
F	4	5	6	7	8	9	10
E	11	12	13	14 Valentine	15	16	17
B	18	19 Pres. Day	20	21 Ash Wednesday	22	23	24
	25	26	27	28	1	2	3
M	4	5	6	7	8	9	10
A	11	12	13	14	15	16	17 St. Patrick
R	18	19	20	21	22	23	24
	25	26	27	28	29	30	31
A	1	2	3 Passover	4	5	6 Good Friday	7
P	8 Easter	9	10	11	12	13	14
R	15	16	17	18	19	20	21
	22	23	24	25	26	27	28
	29	30	1	2	3	4	5
M	6	7	8	9	10	11	12
A	13 Mother's Day	14	15	16	17	18	19
Y	20	21	22	23	24	25	26
	27	28 Memorial Day	29	30	31		

2018	SUN	MON	TUES	WED	THURS	FRI	SAT
						1	2
JUN	3	4	5	6	7	8	9
	10	11	12	13	14	15	16
	17 Father's Day	18	19	20	21	22	23
	24	25	26	27	28	29	30
JUL	1	2	3	4 Independence Day	5	6	7
	8	9	10	11	12	13	14
	15	16	17	18	19	20	21
	22	23	24	25	26	27	28
	29	30	31	1	2	3	4
AUG	5	6	7	8	9	10	11
	12	13	14	15	16	17	18
	19	20	21	22	23	24	25
	26	27	28	29	30	31	1
SEP	2	3 Labor Day	4	5	6	7	8
	9	10	11	12	13 Rosh Hashanah	14	15
	16	17	18	19	20	21	22 Yom Kippur
	23	24	25	26	27	28	29
	30	1	2	3	4	5	6
OCT	7	8 Columbus Day	9	10	11	12	13
	14	15	16	17	18	19	20
	21	22	23	24	25	26	27
	28	29	30	31 Halloween	1	2	3
NOV	4	5	6 Election Day	7	8	9	10
	11	12 Veterans Day Obsv	13	14	15	16	17
	18	19	20	21	22 Thanksgiving	23	24
	25	26	27	28	29	30	1
DEC	2	3	4	5 Hanukkah	6	7	8
	9	10	11	12	13	14	15
	16	17	18	19	20	21	22
	23	24	25 Christmas	26	27	28	29
	30	31 New Yea's Eve					