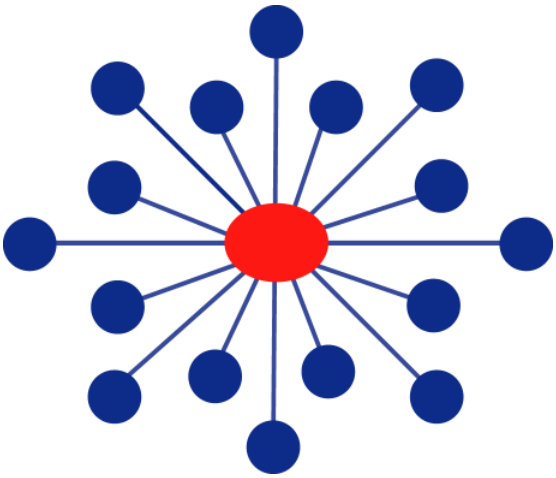
Statistical Analysis Plan (SAP)

CTN-0093: Validation of a Community Pharmacy-Based Prescription Drug Monitoring Program Risk Screening Tool

NCT039369

Document created July 17, 2019



CTN-0093
Statistical Analysis Plan

VALIDATION OF A COMMUNITY PHARMACY-BASED PRESCRIPTION DRUG MONITORING PROGRAM RISK SCREENING TOOL (PHARMSCREEN)

SAP Author: Ziji Yu, PhD

Senior Statistician: Ben Haaland, PhD

University of Utah Statistical Design and Biostatistics Center

June 17, 2019

1.0 STUDY DESIGN AND OBJECTIVES

1.1 STUDY DESIGN

This study is a one group, cross-sectional, validation study. Approximately 1,523 adults prescribed opioid medications will be recruited from one of 15 participating Kroger Pharmacies. Two sources of data will be obtained for this study. The first source is participants will complete self-report assessments via a secure on-line portal regarding contact; demographic; and behavioral and physical health information. The second source is the Narcotic Score (NS), referred to as the "NS metric" herein, which will be provided by Appriss Health.

1.2 STUDY OBJECTIVES

The goal of the study is to validate NS metric as Prescription Drug Monitoring Program-based instrument to discriminate between low, moderate, and high-risk opioid use disorder. The World Health Organization Alcohol, Smoking, and Substance Involvement Screening Test (WHO ASSIST) will be used as the gold standard instrument that defines patient risk levels. No intervention or hypothesis will be tested.

There are 2 statistical Aims in the project:

Aim 1: To assess the overall ability of NS metric to discriminate between opioid use risk levels and to determine cut-point thresholds for NS metric total score to classify patients into low, moderate, and high risk groups.

Aim 2: To validate the NS metric classifier (with selected cut-point threshold values from Aim 1 as a clinical measure to evaluate patients' opioid risk level, with WHO ASSIST as the gold standard to which the NS metric is compared.

Other exploratory statistical Aims include:

- To explore the association between NS metric and other indicators of opioid use risk, such as: the Tobacco, Alcohol, Prescription medication and other Substances (TAPS) Tool; a positive misuse indication from the Prescription Opioid Misuse Index (POMI); and history of opioid overdose.
- To collect validity data on the TAPS tool in a large sample of individuals filling opioid pain medications.

2.0 GENERAL DEFINITIONS

2.1 DEFINITION OF STUDY POPULATION

2.1.1 ALL ENROLLED SUBJECTS POPULATION

The enrolled subjects population (ENR) will contain all participants who provide informed consent for this study and are deemed as meeting the inclusion/exclusion criteria. ENR will be used to summarize subject disposition.

2.1.2 COMPLETER POPULATION

A participant is considered as a completer if all of the participant's electronic Case Report Forms (eCRFs) have been completed, although it is possible that some questions on the eCRFs may not be answered. The completer population will be used as the main analysis population.

2.2 DEFINITION OF ASSESSMENTS

2.2.1 NS METRIC

Appriss Health has developed the NS metric, which uses PDMP data on prescription opioid and benzodiazepine use and aberrant drug behavior (e.g., multiple providers, pharmacies) to compute a score quantifying the extent of the patient's risk in relation to all prescription opioid users. The NS metric is a continuous indicator on a 000-999 scale, with the last digit representing active number of opioid prescriptions (those with ≥9 prescriptions coded as 9) and the first two numbers representing a composite risk score. Higher scores indicate increased risk for adverse opioid-related outcomes (e.g., overdose). See CTN-0093 Protocol Study Assessments section for details on score composition.

2.2.2 WHO ASSIST

The WHO ASSIST opioid use risk domain will be used as the gold standard to which the NS metric will be compared. The WHO ASSIST opioids domain is a continuous indicator on a 0-20 scale, with the 0-0.5 range representing low risk, the 0.5-14.5 range representing moderate risk, and the 14.5-20 range representing high risk. [1] The WHO ASSIST has demonstrated criterion, construct, concurrent, and discriminant validity. [1]

2.2.3 TAPS

The TAPS Tool consists of a 5-item screening for tobacco use, unhealthy alcohol use, prescription medication misuse, and illicit substance use. Positive reports on the screener result in respondents subsequently being asked 9 additional substance specific items. Scoring on the TAPS Tool involves summing respondents' answers to the substance specific questions, resulting in a 0-3 ordinal score for each drug (0=No use in the past 3 months; 1= problem use; 2= higher risk). The TAPS Tool has demonstrated concurrent validity. [2] This measure will be captured in order to provide additional information regarding its psychometric properties compared to the WHO ASSIST.

2.2.4 OTHER MEASUREMENTS

Covariates known to be correlated with opioid use severity will also be captured. These covariate measures will include: opioid medication misuse, pain severity, general health status, depression, overdose frequency history.

A list of study measurements and domains of interest are summarized in **Table 1:**

Table 1. Study Measures			
Domain	**Name**	**No. of Items**	**Min. to Complete**
Opioid use severity	WHO Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): *Opioid Items*[1]	8	≤5
	WHO Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): *Adapted heroin and prescription opioid items*[3*]	16	≤5
	Tobacco, Alcohol, Prescription Medication, and Other Substance Use (TAPS) 1 / 2 Tool: *Prescription drug and prescription opioid items*[2]	2-4	≤1
	Tobacco, Alcohol, Prescription Medication, and Other Substance Use (TAPS) 1 / 2 Tool: *Illicit drug and heroin items*[2]	2-4	≤1
Opioid medication misuse	Prescription Opioid Misuse Index[4]	6	≤5
Opioid overdose	Overdose Experiences, Self and Witnessed—Drug (OESWD) [5]	1	≤1
Non-opioid drug use severity	WHO Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): *Non-opioid drug use items*[1]	8-72	5-15
	Tobacco, Alcohol, Prescription Medication, and Other Substance Use (TAPS) 1 / 2 Tool: *Non-opioid items*[2]	12-25	≤5
Depression	Patient Health Questionnaire-2[6]	2	≤1
Physical Health	Short Form-12: *General health subscale*[7]	1	≤1
	Brief Pain Inventory[8]	8	≤5
Demographics	PhenX demographics: *age, education, gender, race, ethnicity, health insurance, employment, marital status*	10	<5
	Totals:	**76-157**	**40-50**

[*] The adapted heroin and prescription opioid items will not be employed in the a priori analyses.

3.0 STATISTICAL CONSIDERATION

3.1 MULTICENTER STUDIES

This study will be conducted by multiple investigators at multiple Universities; however, these data will be captured in a single REDCap data system. Descriptive statistics on demographic, other characteristics, and clinical measurements of interests (described in Section 2.2) will be presented by pharmacy recruitment site and between site differences will be examined.

3.2 DATA MANAGEMENT

Throughout the data collection phase of the study, the OVN and University of Utah teams will

collaborate to share a cleaned identified datasets with Appriss Health. Using the patient name, contact information, survey timestamp, and store location; Appriss will deterministically match NS metric data with the final dataset and share a merged dataset with OVN and the University of Utah for analyses. It is important to note that given the privacy of the survey response method (i.e., self-report outside of the pharmacy setting), we anticipate participant responses will have less measurement error (e.g., less social desirability bias). Thus, we anticipate our participant responses will more accurately relate to objective PDMP information than interviewer administered or pharmacy-based surveys.

3.3 DISPOSITION

The number and percentages of screened and enrolled subjects as well as subjects who complete all eCRFs will be presented based on ENR.

3.4 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISITCS

Demographic data and other baseline characteristics will be summarized using descriptive statistics for the Completers Population and ENR. Demographic and other baseline characteristics include, but are not limited to:

- Age (years) - calculated relative to date of submitted e-consent
- Education
- Birth gender
- Race
- Ethnicity
- Health Insurance
- Employment
- Marital status

3.5 STATISTICAL METHODS

The Completer Population will be randomly split into training data set and testing data set with a 1:1 ratio, stratified by WHO ASSIST risk levels. The training data set will be used to select cut-points for NS metric risk levels, and testing data set will be used to address statistical Aim 2, validating the selected cut-points. The proposed approach allows us to assess the predictive ability of the NS metric classifier with an independent testing data set and avoid potentially overfitting.[9] All analyses will be performed both overall and within sex specific subpopulations, to examine sex difference in discrimination, selected cut-points, and agreement between measures.

Summary statistics will be provided for the following clinical assessments for the population, by gender subpopulation, by pharmacy recruitment sites.

- NS Metric
- WHO ASSIST
- TAPS
- OESWD
- Prescription Opioid Misuse Index
- PHQ-2
- SF-12
- Brief Pain Inventory

Statistical Aim 1 – To assess the ability of NS metric to discriminate between opioid use risk groups and to determine cut-off points for NS metric total score.

To assess the ability of NS metric to discriminate between opioid use risk groups:

The discriminative validity of an instrument concerns its ability to discriminate between known groups. In the project, the known groups are the low risk, moderate risk, and high risk groups as defined by WHO ASSIST opioid domain (0-0.5 range=low risk, 0.5-14.5 range=moderate risk, and 14.5-20 range=high risk).

A receiver operating characteristics (ROC) analysis[10] will be conducted to 1) assess the overall discriminative ability of NS metric score as a clinical risk assessment instrument, 2) select cut-point threshold values to define low, moderate, and high risk groups with NS metric.

Area under the ROC curve (AUC) values will be used to determine the overall discriminating ability of NS metric. AUC is an overall performance measurement for the classification of an instrument at various threshold settings, and it tells how much the classifier is capable of distinguish between classes. The following scale will be used for the evaluation: <0.70=poor, ≥0.70=fair, ≥0.80 good ≥0.9=excellent.

The study will be powered to the least prevalent but most severe condition among potential patients. Based on the allocation ratio of the national rate of prescription opioid use disorder (POUD) among those prescribed opioid medications in the last year, the prevalence of POUD is 2.1%. A sample of approximately 1,523 (32 positive group) achieves 97.68% power to detect a difference of 0.2 between the area under the ROC curve (AUC) under the null hypothesis of 0.5 and the alternative hypothesis of 0.7 using a two-sided z-test at a significance level of 0.050.

The Completer Population (both training and testing set) will be used to assess the overall discrimination ability of NS metric.

To determine cut-point threshold values for NS total score:

The objective is to select the optimal cut-points threshold for NS metric to define low, moderate and high risk groups. We will use a hybrid framework where we can make use of the data driven methods to support the clinician's final decision making. Only the training data set will be used in this portion of analysis.

In the first step, the solution space (a subset of optimum choices of cut-point threshold values) will be identified with the following grid search approach: for each pair of risk groups, for example, low risk vs moderate risk, calculate the following performance measurements for each possible cut-point threshold value:

- Sensitivity (SE)

$$\text{SE} = \frac{\text{True Positive Rate}}{\text{True Positive Rate} + \text{False Negative Rate}}$$

- Specificity (SP)

$$\text{SP} = \frac{\text{True Negative Rate}}{\text{True Negative Rate} + \text{False Postive Rate}}$$

- Positive Predictive Value (PPV): probability that the risk level is present (based on WHO ASSIST) when the result with NS metric is positive.

- Negative Predictive Value (NPV): probability that the risk level is not present (based on WHO ASSIST) when the test result with NS metric is negative.

- Youden's index:

$$\text{Youden's Index} = \text{SE} + \text{SP} - 1$$

Youden's index, a metric that summarizes both false positive and false negative misclassification rates with equal weights, will be used as primary evaluation criteria to compare classification performance across scenarios with different cut-points threshold values. A solution space, i.e. a set of cut-point threshold values with highest Youden's Index, will be selected. For example, the selected solution space of low risk vs moderate risk cut-point threshold is 10-30, which gives the highest 10% Youden's Index among all possible choices (0-999). See **Table 2** as an example of grid search result:

Table 2: Grid search result for threshold value of moderate vs high risks							
Possible Threshold values for high risk vs moderate risk	Youden's Index	SE	SP	PPV	NPV	Rank of Youden's Index	Flag for solution space
0	xxx	xxx	xxx	xxx	xxx		
…							
511	xxx	xxx	xxx	xxx	xxx	10/1000	Yes
512	xxx	xxx	xxx	xxx	xxx	100/1000	Yes
513	xxx	xxx	xxx	xxx	xxx	200/1000	No
…							
999	xxx	xxx	xxx	xxx	xxx		

In the second step, the final optimal cut-points threshold values will be selected from the solution space, with clinical justification such that the balance between sensitivity and specificity is achieved. For example, a cut-off point of 20 is selected for low risk vs moderate risk group based upon clinician's justification. This framework will integrate clinical knowledge into purely data-driven techniques.

Assuming the acceptable SE and SP are 0.8 for both cutoffs (low vs moderate/high and low/moderate vs high), with a sample size of 750 patients in the training data set, the width of 95% confidence intervals of SE and SP estimations are 0.026 and 0.056 for the classification cutoffs that predict low risk vs moderate/high (prevalence 82.9% vs 17.1%). For predicting low/moderate risk vs high risk (prevalence 2.1% vs 97.9%), the width of confidence intervals for SE and SP are 0.143 and 0.024, respectively. The narrow confidence intervals imply relatively high precision for the SE and SP estimations, which will be used to support clinicians' decision making.

The final selected cut-points threshold values will be validated with an independent testing data set in Aim 2.

Statistical Aim 2 – To validate discriminative ability of NS metric (with the established risk levels of low, moderate, and high) as a clinical measure to evaluate opioid use risk level.

The second primary aim of the project is to evaluate the concurrent validity of the NS metric with the established risk levels of low, moderate, and high. Classification accuracy will be evaluated with an independent testing data set, and WHO ASSIST will be used as the gold standard to define actual risk levels for patients. The agreement level of risk groups defined by NS metric and WHO ASSIST will be descriptively summarized with a confusion matrix, see **Table 3**. Each row of the matrix represents the instances and percentages in a predicted risk level while each column represents the instances in an actual risk level.

Table 3. Confusion matrix				
		Actual risk levels defined by WHO ASSIST		
		Low risk level	Moderate risk level	High risk level
Predicted risk levels defined by NS metric	Low risk level	XXX(XX.XX%)	XX(XX.XX%)	XX(XX.XX%)
	Moderate risk level	XX(XX.XX%)	XX(XX.XX%)	XX(XX.XX%)
	High risk level	XX(XX.XX%)	XX(XX.XX%)	XX(XX.XX%)

Cohen's Kappa Coefficient[11,12] will be used to evaluate agreement between the two metrics. The following scale will be used to assess level of agreement: 0-0.2=slight, 0.21-0.4=fair, 0.41-0.6=moderate, 0.61-0.8=substantial, and 0.81-1=near perfect. We expect the NS metric will demonstrate at least moderate (≥0.40) agreement with the gold-standard risk categories, and a sample size of 750 patients and prevalence of 2.1% (for high risk level) will achieve 99.8% power with a significance level of 0.05 in a test of H0: Kappa ≤ 0.4 vs. H1: Kappa > 0.4. Agreement between NS metric and WHO ASSIST will also be evaluated via Spearman Correlation[13] based upon risk levels (1 for low risk, 2 for moderate risk, and 3 for high risk for both NS metric and WHO ASSIST). The following scale will be used to assess level of agreement: 0-0.3 = low degree, 0.3-0.5 = moderate degree, 0.5-1 = high degree.

An ordinal logistic regression model[14] will be built to estimate the magnitude of association between the metrics, while controlling for other relevant risk factors. In the ordinal logistic regression, WHO ASSIST risk levels will be considered as response variable, and regression covariates will include NS metric, and other relevant indicators known to be associated with opioid-related risk, including: sex, pain, general health status, depression, and substance use history (**Table 1**). The association between WHO ASSIST and NS metric will be estimated by odds ratios, and Wald statistics will be used to examine the association at a 0.05 significance level.

Exploratory Statistical Aim 1: To validate TAPS tool using WHO ASSIST as gold standard.

The TAPS tool is rapidly becoming recognized as a high quality substance use screener for outpatient health care settings. Given the somewhat limited opioid using sample in the main outcomes study (≤5% for prescription opioids; <4% for heroin*), the current study offers an important opportunity to further validate this instrument in a novel outpatient setting, community pharmacy, among a large opioid using population.

The performance of TAPS Tool will be evaluated relative to WHO ASSIST. We plan to compare subjects' responses to each substance of TAPS Tool with their responses to the corresponding domains of WHO ASSIST. The TAPS tool has four domains, and WHO ASSIST has 9 domains. Domains on the ASSIST will be combined with those of the TAPS Tool, **see Table 4** for details:

Table 4: combinations of substances for WHO ASSIST

		TAPS 1 (past 12 months)			
		Tobacco	Alcohol	Illicit Drug	Prescribed Medication
ASSIST (past 3 months)	Tobacco products (cigarettes, chewing tobacco, cigars, etc.)	X			
	Alcoholic beverages (beer, wine, liquor, etc.)		X		
	Cannabis (marijuana, pot, weed, hash, etc.)			X	
	Cocaine (coke, crack, etc.)			X	
	Amphetamine type stimulants (speed, diet pills, ecstasy, methamphetamine, crystal-meth, etc.)			X	X
	Inhalants (nitrous, glue, poppers, etc.)			X	
	Sedatives or sleeping pills (Valium, Rohypnol, Ativan, Xanax, Klonopin, GHB, etc.)				X
	Hallucinogens (LSD, acid, mushrooms, PCP, Special K, etc.)			X	
	Opioids (heroin, morphine, methadone, codeine, buprenorphine, suboxone, Oxycontin, Percocet, Vicodin, etc.			X	X
	Other (K2, Spice, Synthetic spice, Synthetic marijuana)			X	

The level of agreement between TAPS tool with the established risk levels of low, moderate and high on the WHO ASSIST will be evaluated with similar methods as described in Aim 2: ROC analysis will be used to examine concordance between each domain of the TAPS Tool and the corresponding clustered domain of WHO ASSIST. Risk levels of the WHO ASSIST for each clustered domain are defined as follow:

- WHO ASSIST Tobacco Domain:
    - Moderate and High risk >3
    - Low risk ≤ 3
- WHO ASSIST Alcohol Domain:
    - Moderate and High risk >10
    - Low risk ≤ 10

- WHO ASSIST Illicit Drug Domain:
    - Moderate and High risk >3
    - Low risk ≤ 3
- WHO ASSIST Prescribed Medication Domain:
    - Moderate and High risk >3
    - Low risk ≤ 3.

For each of these binary risk categories, Area under the ROC curve (AUC) values will be used to determine the overall discriminating ability of TAPS tool on each of the 4 domains.

The confusion table will be provided (**see Table 5 example of alcohol domain Confusion Table**) and the following statistical measurements of agreement, all with their respective 95% confidence intervals will be provided: Spearman correlation coefficient, Goodman and Kruskal's Gamma, and Kendall rank correlation coefficient. A logistic regression model will be built to estimate the magnitude of association between the metrics, while controlling for other relevant risk factors.

Table 5. Confusion matrix for Alcohol Domain

		TAPS Tool observation				
		0	1	2	3	4
Predicted risk levels defined by WHO ASSIST	Low risk level	XX (XX.XX%)	XX (XX.XX%)	XX (XX.XX%)	XX (XX.XX%)	XX (XX.XX%)
	High/ moderate risk level	XX (XX.XX%)	XX (XX.XX%)	XX (XX.XX%)	XX (XX.XX%)	XX (XX.XX%)

Exploratory Statistical Aim 2: To assess the association between the NS metric and other measurements of opioid misuse.

In this exploratory aim, we will assess the relationship between the NS metric with the established risk levels of low, moderate and high risk and 1) ordinal prescription and heroin use scores from the TAPS Tool (prescribed medication and heroin domains), 2) a positive misuse indication from (POMI), and 3) overdose frequency history.

Risk levels will be defined as follow:

- TAPS Tool prescribed medication domain:
    - High risk≥2;
    - Low risk<2.

- POMI:
    - High risk ≥2;
    - Low risk <2.
- Overdose Frequency History - OESWD domain:
    - Any Overdose >0;
    - No Overdose = 0.

We will likewise employ the above described statistical analyses for Exploratory Aim 1 and assess the association and agreement level between NS metric and the TAPS Tool, POMI, and Overdose Frequency History with the:

- Confusion table;
- AUC of ROC analysis;
- Spearman Correlation Coefficient;
- Goodman and Kruskal's Gamma;
- Kendall rank correlation coefficient;
- Ordinal Logit Regression.

Exploratory Statistical Aim 3: To validate definition of NS metric.

In this section, we will validate the definition of NS metric and test if the third digit of NS metric (number of opioid prescription) is informative. We will examine an alternative definition of NS metric, named as NS metric v2, which consists of only the first two numbers of original NS metric, which ranges 00-99. Descriptive analyses will be conducted and correlation between NS metric and NS metric v2 will be examined with Spearman Correlation Coefficients. We will repeat the planned training/validation analyses in the primary analysis and compare the discriminative ability of NS metric v2 vs. NS metric. We expect the predictive power of NS metric v2 is lower than that of NS metric, which implies the 3rd digit in NS metric original definition is predictive. We will also repeat the training/validation analysis within each subgroup of patients with same numbers of opioid prescriptions. The overall discriminate ability for NS metric 2, described by average misclassification rate across subgroups, will be compared with primary analyses results. It is expected that the prediction accuracy is comparable, which implies the "opioid prescription number" information is well represented by the third digit of original NS metric definition.

3.6 EXAMINATION OF SUBGROUPS

Subgroup analyses will be conducted on the primary and exploratory aims for the following groups:

- Gender: male and female.

For the subgroup analyses, all completers in each subpopulation will be randomly split into training data set and testing data set with a 1:1 ratio, stratified by WHO ASSIST risk levels. The cut-off threshold values will be selected using similar methods as described in Aim 1 for each subgroup and validated using similar methods as described in Aim 2, with the testing set for the corresponding subgroup. We will compare the selected threshold values and their discriminative ability between subgroups, see **Table 6** for details. *It should be noted* that the study is *not* designed to detect certain agreement level between metrics with high statistical power within any subgroup.

Table 6. Summary table that compares discrimination ability of NS metrics with established cut-off threshold between 2 gender subgroups, with WHO ASSIST as gold standard.

		Female	Male	Overall
AUC of ROC	Low risk vs Moderate risk			
	Moderate risk vs high risk			
NS Metric cut-off values	Low risk vs Moderate risk			
	Moderate risk vs high risk			
Validation of selected cut-off values				
Kappa coefficient				
Spearman Correlation Coefficient				
Goodman and Kruskal's Gamma				
Kendall rank correlation coefficient				
Odds ratio from ordinal logit regression				

3.7 MISSING DATA

The completer population, defined as participants who complete all opioid outcome score contributing items on the WHO ASSIST and TAPS tool, will be used for the main analysis. Completers also must have a NS metric score. As a result of the one-time cross-sectional nature of this study, participant attrition is not a concern. Rubin's multiple imputation approach will be used to impute missing covariates but missing outcome data will not be imputed.[15]

3.8 OUTPUT PRESENTATION

APPENDIX 1 shows conventions for presentation of data in outputs. The tables, figures, and listings (TFL) shells provided with this SAP describe the presentations for this study and therefore the format and content of the summary tables, figures and listings to be provided by SDBC Statisticians.

3.9 SOFTWARE

All analyses performed by the SDBC will use R 3.4.4 for windows software.

APPENDIX 1 TABLES, FIGURE AND LISTING SHELLS

Proposed Tables, Figures and Listings			
Section	**Table/Figure #**	**Analysis Population**	**Title**
Enrollment and Participant Status	Table 1.1	ENR	Summary of Subject Disposition
	Table 1.2	ENR	Summary of Subject Disposition by Gender
	Table 1.3	CP	Summary of Training/Testing Random Split, Overall and by Gender
Baseline Demographics	Table 2.1	CP	Summary of Baseline Demographics
	Table 2.2	CP	Summary of Baseline Demographics by Gender
	Table 2.3	CP	Summary of Baseline Demographics by Investigation Site
Summary of Clinical Assessment	Table 3.1	CP	Descriptive summary of WHO ASSIST
	Table 3.2	CP	Descriptive summary of WHO ASSIST, by Gender
	Table 3.3	CP	Descriptive summary of WHO ASSIST, by Investigation site
	Table 3.4	CP	Descriptive summary of NS Metric
	Table 3.5	CP	Descriptive summary of NS Metric, by Gender
	Table 3.6	CP	Descriptive summary of NS Metric, by Investigation site
	Table 3.7	CP	Descriptive summary of TAPS Tool
	Table 3.8	CP	Descriptive summary of TAPS Tool, by Gender
	Table 3.9	CP	Descriptive summary of TAPS Tool, by Investigation site
	Table 3.10	CP	Descriptive summary of POMI Tool
	Table 3.11	CP	Descriptive summary of POMI Tool, by Gender
	Table 3.12	CP	Descriptive summary of POMI Tool, by Investigation site
	Table 3.13	CP	Descriptive summary of OESWD
	Table 3.14	CP	Descriptive summary of OESWD, by Gender
	Table 3.15	CP	Descriptive summary of OESWD, by Investigation site
	Table 3.16	CP	Descriptive summary of Prescription Opioid Misuse Index
	Table 3.17	CP	Descriptive summary of Prescription Opioid Misuse Index, by Gender
	Table 3.18	CP	Descriptive summary of Prescription Opioid Misuse Index, by Investigation site
	Table 3.19	CP	Descriptive summary of PHQ-2
	Table 3.20	CP	Descriptive summary of PHQ-2, by Gender
	Table 3.21	CP	Descriptive summary of PHQ-2, by Investigation site
	Table 3.22	CP	Descriptive summary of SF12

Section	Table/Figure #	Analysis Population	Title
	Table 3.23	CP	Descriptive summary of SF12, by Gender
	Table 3.24	CP	Descriptive summary of SF12, by Investigation site
	Table 3.25	CP	Descriptive summary of Brief Pain Inventory
	Table 3.26	CP	Descriptive summary of Brief Pain Inventory, by Gender
	Table 3.27	CP	Descriptive summary of Brief Pain Inventory, by Investigation site
Results for Aim 1	Figure 3.1	CP	NS metric ROC curve for discriminating high risk vs moderate risk opioid misuse patients defined by WHO ASSIST opioid items.
	Figure 3.2	CP	NS metric ROC curve for discriminating moderate risk vs low risk opioid misuse patients defined by WHO ASSIST opioid items.
	Figure 3.3	CP	NS metric ROC curve for discriminating high risk vs moderate risk opioid misuse patients defined by WHO ASSIST opioid items, by gender
	Figure 3.4	CP	NS metric ROC curve for discriminating moderate risk vs low risk opioid misuse patients defined by WHO ASSIST opioid items, by gender
	Table 3.5	CP	Summary of AUC of ROC curve
	Table 3.6	CP	Summary of AUC of ROC curve, by gender
	Table 3.7	CP - Training	Summary of grid search: ROC analysis results for all possible cut-off threshold values of high risk vs moderate risk with NS metric.
	Table 3.8	CP - Training	Summary of grid search: ROC analysis results for all possible cut-off threshold values of moderate risk vs low risk with NS metric.
	Table 3.9	CP - Training	Summary of grid search: ROC analysis results for all possible cut-off threshold values of high risk vs moderate risk with NS metric by gender
	Table 3.10	CP - Training	Summary of grid search: ROC analysis results for all possible cut-off threshold values of moderate risk vs low risk with NS metric by gender
	Figure 3.11	CP - Training	NS metric ROC curve for discriminating high risk vs moderate risk opioid misuse patients defined by WHO ASSIST opioid items, with training data set.
	Figure 3.12	CP - Training	NS metric ROC curve for discriminating moderate risk vs low risk opioid misuse patients defined by WHO ASSIST opioid items, with training data set.
	Figure 3.13	CP -	NS metric ROC curve for discriminating high risk

Section	Table/Figure #	Analysis Population	Title
		Training	vs moderate risk opioid misuse patients defined by WHO ASSIST opioid items, with training data set, by gender.
	Figure 3.14	CP - Training	NS metric ROC curve for discriminating moderate risk vs low risk opioid misuse patients defined by WHO ASSIST opioid items, with training data set, by gender.
	Figure 3.15	CP- Training	Plot of Kappa Coefficient for NS metric vs WHO Assist, for all possible threshold values in selected solution space
	Figure 3.16	CP - Training	Plot of Kappa Coefficient for NS metric vs WHO Assist, for all possible threshold values in selected solution space, by gender
	Figure 3.17	CP - Training	Plot of Spearman Correlation Coefficient for NS metric vs WHO Assist, for all possible threshold values in selected solution space
	Figure 3.18	CP - Training	Plot of Spearman Correlation Coefficient NS metric vs WHO Assist, for all possible threshold values in selected solution space, by gender
	Figure 3.19	CP - Training	Plot of Specificity and Sensitivity, for all possible high vs moderate risk threshold in solution space
	Figure 3.20	CP- Training	Plot of Specificity and Sensitivity, for all possible moderate risk vs low risk threshold in solution space
	Figure 3.21	CP - Training	Plot of Specificity and Sensitivity, for all possible high vs moderate risk threshold in solution space, by gender
	Figure 3.22	CP- Training	Plot of Specificity and Sensitivity, for all possible moderate risk vs low risk threshold in solution space, by gender
	Figure 3.23	CP - Training	Plot of Positive Predictive Value and Negative Predictive Value, for all possible high vs moderate risk threshold in solution space
	Figure 3.24	CP- Training	Plot of Positive Predictive Value and Negative Predictive Value, for all possible moderate vs mild risk threshold in solution space
	Figure 3.25	CP- Training	Plot of Positive Predictive Value and Negative Predictive Value, for all possible high vs moderate risk threshold in solution space, by gender
	Figure 3.26	CP - Training	Plot of Positive Predictive Value and Negative Predictive Value, for all possible moderate vs mild risk threshold in solution space, by gender
	Table 3.27	CP -	Summary of ROC analysis result and Correlation

Section	Table/Figure #	Analysis Population	Title
		Training	Analysis result for the selected cut-off threshold value of NS metric, high vs moderate risk classification and moderate vs low risk classifcation
Results for Aim 2	Table 4.1	CP - Testing	Confusion Table for NS metric with established cut-off threshold vs WHO ASSIST
	Table 4.2	CP - Testing	Confusion Table for NS metric with established cut-off threshold vs WHO ASSIST, by gender
	Table 4.3	CP - Testing	Summary of Kappa Coefficient and Correlation Analysis Results, WHO ASSIST vs NS metric (with established cut-off value from Aim 1)
	Table 4.4	CP - Testing	Summary of Kappa Coefficient and Correlation Analysis Results, WHO ASSIST vs NS metric, by gender (with established cut-off value from Aim 1)
	Table 4.5	CP - Testing	Summary of ordinal logit regression results (with established cut-off value from Aim 1)
	Table 4.8	CP - Testing	Summary of ordinal logit regression results, by gender (with established cut-off value from Aim 1)
Results for Exploratory Aim 1	Table 5.1	CP	TAPS Tool ROC curve for discriminating high risk vs low risk groups defined by WHO ASSIST, by domains (Tobacco, Alcohol, Illicit Drug, Prescribed Medication)
	Table 5.2	CP	TAPS Tool ROC curve for discriminating high risk vs low risk groups defined by WHO ASSIST, by domains (Tobacco, Alcohol, Illicit Drug, Prescribed Medication), by gender
	Table 5.3	CP	Summary of AUC of ROC curve for each domain
	Table 5.4	CP	Summary of AUC of ROC curve for each domain by gender
	Table 5.5	CP	Confusion Table for TAP Tool vs WHO ASSIST, for each domain
	Table 5.6	CP	Confusion Table for TAP Tool vs WHO ASSIST, for each domain, by gender
	Table 5.7	CP	Summary of Kappa Coefficient and Correlation Analysis Results, WHO ASSIST vs TAP Tools for each domain
	Table 5.8	CP	Summary of Kappa Coefficient and Correlation Analysis Results, WHO ASSIST vs TAP Tools for each domain, by gender
	Table 5.9	CP	Summary of ordinal logit regression results for each domain
	Table 5.10	CP	Summary of ordinal logit regression results for

Section	Table/Figure #	Analysis Population	Title
			each domain, by gender
Results for Exploratory Aim 2 – explore association between NS metric vs TAP Tools	Table 6.1.1	CP	Confusion Table for NS metric with established cut-off threshold vs TAPS Prescribed Medication Domain
	Table 6.1.2	CP	Confusion Table for NS metric with established cut-off threshold vs TAPS Prescribed Medication Domain, by gender
	Figure 6.1.3	CP	NS metric ROC curve for discriminating high risk vs low risk opioid misuse patients (defined by TAPS)
	Figure 6.1.4	CP	NS metric ROC curve for discriminating high risk vs low risk opioid misuse patients (defined by TAPS), by gender
	Table 6.1.5	CP	Summary of AUC of ROC curve, for NS metric vs TAPS
	Table 6.1.6	CP	Summary of AUC of ROC curve, for NS metric vs TAPS by gender
	Table 6.1.7	CP	Summary of Kappa Coefficient and Correlation Analysis Results, NS metric vs TAP Tools
	Table 6.1.8	CP	Summary of Kappa Coefficient and Correlation Analysis Results, NS metric vs TAP Tools, by gender
	Table 6.1.9	CP	Summary of ordinal logit regression results, NS metric vs TAP Tools
	Table 6.1.10	CP	Summary of ordinal logit regression results, NS metric vs TAP Tools, by gender
Results for Exploratory Aim 2 – explore association between NS metric vs POMI Tools	Table 6.2.1	CP	Confusion Table for NS metric with established cut-off threshold vs POMI
	Table 6.2.2	CP	Confusion Table for NS metric with established cut-off threshold vs POMI, by gender
	Figure 6.2.3	CP	NS metric ROC curve for discriminating high risk vs low risk opioid misuse patients (defined by POMI)
	Figure 6.2.4	CP	NS metric ROC curve for discriminating high risk vs low risk opioid misuse patients (defined by POMI), by gender
	Table 6.2.5	CP	Summary of AUC of ROC curve, for NS metric vs POMI
	Table 6.2.6	CP	Summary of AUC of ROC curve, for NS metric vs POMI by gender
	Table 6.2.7	CP	Summary of Kappa Coefficient and Correlation Analysis Results, NS metric vs POMI
	Table 6.2.8	CP	Summary of Kappa Coefficient and Correlation Analysis Results, NS metric vs POMI, by gender

Section	Table/Figure #	Analysis Population	Title
	Table 6.2.9	CP	Summary of ordinal logit regression results, NS metric vs POMI
	Table 6.2.10	CP	Summary of ordinal logit regression results, NS metric vs POMI, by gender
Results for Exploratory Aim 2 – explore association between NS metric vs OESWD	Table 6.3.1	CP	Confusion Table for NS metric with established cut-off threshold vs OESWD
	Table 6.3.2	CP	Confusion Table for NS metric with established cut-off threshold vs OESWD, by gender
	Figure 6.3.3	CP	NS metric ROC curve for discriminating high risk vs low risk opioid misuse patients (defined by OESWD)
	Figure 6.3.4	CP	NS metric ROC curve for discriminating high risk vs low risk opioid misuse patients (defined by OESWD), by gender
	Table 6.3.5	CP	Summary of AUC of ROC curve, for NS metric vs OESWD
	Table 6.3.6	CP	Summary of AUC of ROC curve, for NS metric vs OESWD by gender
	Table 6.3.7	CP	Summary of Kappa Coefficient and Correlation Analysis Results, NS metric vs OESWD
	Table 6.3.8	CP	Summary of Kappa Coefficient and Correlation Analysis Results, NS metric vs OESWD, by gender
	Table 6.3.9	CP	Summary of ordinal logit regression results, NS metric vs OESWD
	Table 6.3.10	CP	Summary of ordinal logit regression results, NS metric vs OESWD, by gender
Results for Subgroup Comparison	Table 7.1	CP	Summary table that compares selected cut-off threshold values with NS metric between 2 gender subgroups.
	Table 7.2	CP	Summary table that compares discrimination ability of NS metrics with established cut-off threshold between 2 gender subgroups, with WHO ASSIST as gold standard.
	Table 7.3	CP	Summary table that compares discrimination ability of TAPS tool between 2 gender subgroups, with WHO ASSIST as gold standard.
	Table 7.3	CP	Summary table that compares between metrics association of NS metrics vs 1)WHO ASSIST 2) TAPS 3) OESWD 4) POMI, between 2 gender subgroups
Protocol Deviations	Table 8.1	CP	Summary of Protocol Deviations in Study Completers
	Listing 8.2	CP	Listing of Protocol Deviations in Study

Section	Table/Figure #	Analysis Population	Title
			Completers

REFERENCES

1. Humeniuk R, Ali R. *Validation of the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) and Pilot Brief Intervention: A Technical Report of Phase II Findings of the WHO ASSIST Project.* Geneva: WHO;2006.
2. McNeely J, Wu LT, Subramaniam G, et al. 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.* 2016;165(10):690-699.
3. McNeely J, Strauss SM, Rotrosen J, Ramautar A, Gourevitch MN. 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.* 2016;111(2):233-244.
4. Knisely JS, Wunsch MJ, Cropsey KL, Campbell ED. Prescription Opioid Misuse Index: A brief questionnaire to assess misuse. *Journal of Substance Abuse Treatment.* 2008;35(4):380-386.
5. Fernandez AC, Bush C, Bonar EE, Blow FC, Walton MA, Bohnert ASB. Alcohol and Drug Overdose and the Influence of Pain Conditions in an Addiction Treatment Sample. *Journal of addiction medicine.* 2019;13(1):61-68.
6. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener. *Medical Care.* 2003;41(11):1284-1292.
7. Luo X, George ML, Kakouras I, et al. Reliability, validity, and responsiveness of the short form 12-item survey (SF-12) in patients with back pain. *Spine.* 2003;28(15):1739-1745.
8. Keller S, Bann CM, Dodd SL, Schein J, Mendoza TR, Cleeland CS. Validity of the brief pain inventory for use in documenting the outcomes of patients with noncancer pain. *Clin J Pain.* 2014;20(5):309-319.
9. Cawley G, Talbot N. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *J Mach Learn Res.* 2010;11:2079-2107.
10. Youngstrom EA. A Primer on Receiver Operating Characteristic Analysis and Diagnostic Efficiency Statistics for Pediatric Psychology: We Are Ready to ROC. *Journal of Pediatric Psychology.* 2014;39(2):204-221.
11. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement.* 1960;20(1):37-46.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
13. Kendall MG. *Rank correlation methods* London: Griffin; 1970.
14. McCullagh P. Regression Models for Ordinal Data. *Journal of the Royal Statistical Society: Series B (Methodological).* 1980;42(2):109-127.
15. Rubin D. *Multiple imputation for nonresponse in surveys* Vol 81. Hoboken, NJ: John Wiley & Sons; 2004.