An Open-Label, Parallel, Randomized Study to Evaluate the Performance of Needle Placements for Diagnostic and Therapeutic Neuraxial Procedures, Using a Handheld Tactile Imaging-based Method Versus Palpation

NCT04796935

Document Type: Statistical Analysis Plan

Document Date: December 22, 2020

IntuiTap, VerTouch

An Open-Label, Parallel, Randomized Study to Evaluate the Performance of Needle Placements for Diagnostic and Therapeutic Neuraxial Procedures, Using a Handheld Tactile Imaging-based Method Versus Palpation

Prepared by

Stat One, LLC

Document Revision History:

Version	Date	Content/Modifications
01	17 November 2020	Initial draft
02	22 December 2020	Updated based on protocol update incorporating FDA comments

Statistical Analysis Plan
Approval
Signature Page

	Signature of Approval:	Date:
Author: John Davidson, MS Biostatistician Stat One, LLC	John Davidson	12-22-2020
Review: George DeMuth, MS Biostatistician Stat One, LLC	George DeMuth	22DEC2020
Approval: Nicole Moskowitz, MS CTO IntuiTap Medical, Inc	DocuSigned by: Nicole C Moskowitz 1B8CBFE8A646422...	12/22/2020

Table of Contents

1 Abbreviations

Expand as needed

AE	Adverse Event
ATC	Anatomical Therapeutic Chemical
CI	Confidence Interval
CMH	Cochran-Mantel-Haenzel
ITT	Intent to Treat
LOCF	Last Observation Carried Forward
MAR	Missing at Random
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent to Treat
NDA	New Drug Application
PMA	Premarket Approval
PT	Preferred Term
SAE(s)	Serious Adverse Event(s)
SAS®	Statistical Analysis Software®
SOC	System Organ Class

2 Introduction

2.1 Background

Neuraxial procedures, in which a needle is inserted into the spinal canal through a gap in the vertebrae, are performed at a rate of nearly 13 million per year in the US. The current standard of care, involving manual palpation of the patient's back in order to detect the spinous processes and estimate the location of the interspinous needle insertion site, can be inaccurate. This inaccuracy often leads to multiple insertion attempts, thus increasing patient pain, the potential for complications, unpredictable procedure times, and overall patient dissatisfaction. Patients with a high body mass index (BMI) further exacerbate this inaccuracy which has led to an increase in procedure failures as obesity rates have climbed in recent years. Challenging cases are often referred to radiology for fluoroscopic guidance or an ultrasound may be used to help visualize the vertebrae. Each of these solutions present their own issues. Fluoroscopy exposes patients and providers to unnecessary radiation, can rarely be used in perioperative settings, and is associated with increased costs while ultrasound modality requires significant training on how to interpret its output, and is cumbersome, requiring a gel medium and manipulation of a needle in combination with an unstable probe. The VerTouch™ device seeks to overcome these shortcomings by employing scanning-based tactile imaging in order to detect spinal landmarks, analogous to those sensed during palpation, thus increasing the success of needle insertion while also requiring minimal training.

This analysis plan describes an open-label, parallel, randomized study comparison between the VerTouch versus the conventional palpation technique. Subjects enrolled in the study will be patients undergoing neuraxial procedures evenly represented across emergency medicine, neurology, and anesthesiology settings. Enrolled subjects will be randomized into a 1:1 ratio between IntuiTap and the standard palpitation treatment groups

Endpoints of interest in the analysis include the primary efficacy endpoint mean number of insertion attempts, and the primary safety endpoints of incidence of PDPH, unintended dural puncture, paresthesia during needle insertion, and traumatic tap. Other data collected may be analyzed as secondary or tertiary analyses.

2.2 Changes to Planned Analyses

This is the initial version of the statistical analysis plan.

3 Study Objectives

The primary objective of this study is to establish the superiority of VerTouch versus the conventional palpation technique for the number of needle insertion attempts required in diagnostic and therapeutic neuraxial procedures.

4 Investigational Plan and Study Design

Subjects will be randomized 1:1 into two groups via the block randomization method with a list held in each test setting. Standard landmarking techniques will be used in both groups to identify the procedure area. For subjects randomized to the tactile-imaging group (T), VerTouch will be used to identify an interspinous space and to place a marker, introducer, or needle. For subjects randomized to the palpation group (P), the palpation-landmarking method will be used. After marker, introducer, or needle placement, the procedure will continue in the usual manner for subjects in both groups.

The study observer within each setting will record the time to identify an insertion site; time to place the marker, introducer, or needle; number of insertions, re-directions, and bone contacts (counted until confirmation of spinal canal access can be assessed); incidence of procedure success; subject's level of discomfort during landmarking; and provider's level of confidence with the identified insertion site. Total procedure time, incidence of traumatic taps, and incidence of referrals to radiology or pain management will be recorded in all settings. In anesthesiology, incidences of paresthesia, unintended dural puncture, and conversion from spinal to epidural anesthesia, will be collected for exploratory analysis, where applicable. Any usability issues or device malfunctions observed during the clinical study will also be documented.

EM, neurology, and anesthesiology settings will be approximately evenly represented in the study, with at least 24 subjects enrolled in each. To adequately represent patients undergoing obstetric and orthopedic procedures, enrolled anesthesiology subjects will be approximately evenly split between these sub-settings. The PM setting will not be included in the study, as there are limitations, particularly with respect to reimbursement, for ESIs performed with palpation or other non-fluoroscopic techniques. However, PM providers will be represented in IntuiTap's summative usability study and undergo the same anesthesiology training as those clinical study investigators specialized in obstetrics and orthopedics.

Investigators will form a representative sample of typical neuraxial procedure providers, including MDs (residents, fellows, and attendings), PAs, and CRNAs, where applicable. Residents must be in their second post-graduate year and above, having performed at least 5 neuraxial procedures in the past 12 months, and are considered trainees. Apart from their level of training, investigators should have a representative range of relevant experience, indicated by the approximate number of neuraxial procedures they have performed in the past year. To ensure adequate user variability, a single investigator should perform no more than 20% of procedures in his/her setting. Any provider type not included in the study will have been assessed in formative and/or summative human factors testing.

5 Sample Size Justification

It is expected that approximately 120 subjects will be screened to meet a target enrollment goal of 96 subjects (48 in each group) across 3 investigational sites. The superiority hypothesis for the number of insertion attempts with IntuiTap (μ_1) and palpation (μ_2) is:

$$H_0: \mu_1 \geq \mu_2 \text{ vs } H_1: \mu_1 < \mu_2$$

The sample size was selected based on historical information and previous clinical trial data collected with the VerTouch device. A literature survey identified eight (8) papers where the mean and standard deviation (SD) for insertions for palpation were identified[1-8]. The definitions appear similarly to the proposed endpoint. The mean attempts ranged from 1.3 to 3.3 and SD ranged from 0.0 to 6.9. A weighted average of the means and standard deviations provided an average of mean of 2.3 and SD of 2.8[9]. Based on the spread of individual results, a mean of 2.1 and SD of 1.5 was considered a reference for the power analysis.

A sample of 81 attempts was available for the VerTouch device across several iterations of the device. Of the attempts, 90% were successes on the first attempt and the mean number of attempts was 1.1 and SD was 0.6.

In order to assess the sample size, given the truncated nature of the distribution and the planned use of a bootstrap analysis, a simulation study was performed. Random vectors were built by taking the ceiling of absolute values of a mixture for two normal distributions to build long-tailed distributions with means and standard deviations similar to those above. Analyses were done using a bootstrap analysis of the difference in means, Wilcoxon Test, T-test with unequal means, and a T-test of log-transformed data. For a sample size of 24 subjects per group, the bootstrap analysis, Wilcoxon Test, and T-test of log-transformed data had at least 92% power while a T-test of the observed counts had 89% power.

In addition, an estimate of the secondary endpoint of first attempt success of approximately 50% was obtained in the simulation exercise for palpation. The planned sample size of 48 per group provides 90% power when compared to a first attempt success rate of 81% in the VerTouch group. Hence, the study is also expected to have adequate power for the primary analysis and the secondary endpoint of success on the first insertion.

6 Visit Schedule and Visit Windows

At the time of treatment, data will be collected on concomitant medication; pain and/or anti-anxiety medication(s) given for the procedure; provider level of training, number of years in practice, experience, and specialty; procedure setting; and type of and reason for neuraxial procedure. The subject will be treated, either with the device or the control. The designated observer will also be responsible for counting needle insertions and redirections. Post treatment, the subject will be monitored for adverse events, adverse device effects or other reportable observations to ensure safety, entering relevant data into the eCRF.

A follow-up visit will be conducted at 3 ± 2 days following treatment. The presence of PDPH and potential adverse events will be assessed in all subjects. At the conclusion of the follow-up visit, a Subject Study Exit CRF will be completed. Adverse events identified in the study will be followed to resolution.

Assessment	Visit 1	Visit 2

	Screening	Treatment	Post-Treatment	Follow-up
Demographics	X			
Medical History	X			
Physical Exam (Height & Weight)	X			
Inclusion/Exclusion Criteria	X			
Informed Consent	X			
Randomization	X			
Concomitant Medications	X	X		
Procedure Completion		X		
Discomfort Assessment			X	
Provider Satisfaction Assessment			X	
PDPH			X	
Adverse Event Assessment		X		X
Adverse Device Effect Assessment		X		X

7 Study Populations

The study populations are defined as follows:

- Intent-to-treat population (ITT): All randomized subjects according to their randomized group.
- Per-protocol population (PP): All subjects who received their randomized treatment, did not have an inclusion or exclusion criteria violation, had a procedure success, and did not have a major protocol deviation.

The primary analysis and all study results will be analyzed using the ITT population. The primary analysis will be repeated using the PP population. Subjects who meet the discontinuation criteria will be discontinued from the study.

8 Statistical Methods

8.1 General Reporting Conventions

Summary tables will be generated to support the analysis using the subject treatment as a group. The numeric descriptive statistics include the n (observed data), mean, standard deviation (SD), median, minimum value, and maximum value. Categorical summaries will show the number and percent of subjects in the levels associated with the variable summarized and the analysis performed. Percentages will be calculated based on the total number of subjects with non-missing data for the assessment. The denominator should be included on each table to indicate the set of subjects included in each analysis.

Statistical analyses will be performed to assess the homogeneity of the study populations, evaluate the primary hypothesis, and evaluate the impact of covariates on the performance of the VerTouch device. Standard statistical tests used for hypothesis testing are the t-test for numeric measures and the likelihood ratio chi-square test for categorical variables. For categorical variables with small cell counts (less than 5), Fisher's exact test will be used. For inferential analyses, unless otherwise indicated, a two-sided p-value of less than or equal to 0.05 will be considered statistically significant. All confidence intervals (CIs) will be two-sided 95% intervals. Note that p-values reported for tertiary and exploratory endpoints will be used for reporting purposes but are not to be considered for use in product labeling.

8.2 Adjustments for Interim Analyses and Multiplicity of Endpoints

The study has a single primary endpoint needed for the superiority analysis. If that analysis reaches statistical significance, then the following additional hypothesis of the secondary variables will be considered sequentially such that if one analysis is statistically significant, then the next hypothesis can be evaluated to control for the Type I error for these analyses:

Additional Hypothesis 1: The null hypothesis that the binary rate of first-attempt success is the same in both groups will be analyzed using a likelihood ratio chis-square test (H_0: $p_1 = p_2$ vs H_1: $p_1 \neq p_2$ where p_1 is the success rate in VerTouch subjects and p_2 is the success rate in control subjects). The results will be considered statistically significant if the two-sided p-value is less than 0.05 and the VerTouch success rate is higher than the control rate).

Additional Hypothesis 2: The null hypothesis that the mean number of redirections is the higher in the VerTouch subjects compared to the control group using a T-test with unequal variance assumption (H_0: $\mu_1 \geq \mu_2$ vs H_1: $\mu_1 < \mu_2$ where μ_1 is the mean in VerTouch subjects and μ_2 is the mean in control subjects). The results will be considered statistically significant if the one-sided p-value is less than 0.025.

Additional Hypothesis 3: The null hypothesis that the mean number of passes is the higher in the VerTouch subjects compared to the control group using a T-test with unequal variance assumption (H_0: $\mu_1 \geq \mu_2$ vs H_1: $\mu_1 < \mu_2$ where μ_1 is the mean in VerTouch subjects and μ_2 is the mean in control subjects). The results will be considered statistically significant if the one-sided p-value is less than 0.025.

Additional Hypothesis 4: The null hypothesis that the binary rate of first-pass success is the same in both groups will be analyzed using a likelihood ratio chis-square test (H_0: $p_1 = p_2$ vs H_1: $p_1 \neq p_2$ where p_1 is the success rate in VerTouch subjects and p_2 is the success rate in control subjects). The results will be considered statistically significant if the two-sided p-value is less than 0.05 and the VerTouch success rate is higher than the control rate).

Additional Hypothesis 5: The null hypothesis that the mean subject discomfort rated during landmarking from 0 to 10 is higher in the VerTouch subjects compared to the control group using a T-test with unequal variance assumption (H_0: $\mu_1 \geq \mu_2$ vs H_1: $\mu_1 < \mu_2$ where μ_1 is the mean in VerTouch subjects and μ_2 is the mean in control subjects). The results will be considered statistically significant if the one-sided p-value is less than 0.025.

Additional Hypothesis 6: The null hypothesis that the mean provider confidence with the identified insertion site rated 1 to 5 is lower in the VerTouch subjects compared to the control group using a T-test with unequal variance assumption (H_0: $\mu_1 \leq \mu_2$ vs H_1: $\mu_1 > \mu_2$ where μ_1 is the

mean in VerTouch subjects and μ_2 is the mean in control subjects). The results will be considered statistically significant if the one-sided p-value is less than 0.025.

Additional Hypothesis 7: The null hypothesis that the binary rate of procedure success is the same in both groups will be analyzed using a likelihood ratio chis-square test (H_0: $p_1 = p_2$ vs H_1: $p_1 \neq p_2$ where p_1 is the success rate in VerTouch subjects and p_2 is the success rate in control subjects). The results will be considered statistically significant if the two-sided p-value is less than 0.05 and the VerTouch success rate is higher than the control rate).

8.3 Evaluations and Statistical Analyses

8.3.1 Subject Disposition

Study accountability information will be summarized through:

- The total number of patients randomized
- The number of ITT and PP subjects in each treatment group.
- The number and percent of subjects with completing the baseline information, the procedure, and follow-up visit
- The number and percentage of subjects completing the study, discontinuing in the study, and by any reason for discontinuation
- The number and percentage of subjects with any protocol deviation, total number of deviations, and deviations by type
- The number and percentage of subjects with a device deviation and the type of deviations reported.

8.3.2 Demographics and Baseline Characteristics

Descriptive summaries will be provided for demographic and medical history information for all Intent-to-Treat subjects. Such summaries may include, age, sex, race, pain medications, and other relevant medical history recorded.

8.3.3 Primary Efficacy Endpoint

The primary efficacy endpoint of the study is designed to evaluate whether the use of the VerTouch device is able to reduce the number of insertion attempts to successfully access the spinal canal in a spinal puncture as compared to the conventional palpation technique. Attempts are counted until confirmation of spinal canal access can be assessed.

8.3.3.1 Primary Efficacy Hypothesis

The primary efficacy superiority hypothesis for the average number of insertion attempts to successfully access the spinal canal in a spinal puncture as compared to the conventional palpation technique. The null hypothesis for VerTouch ($\mu 1$) and palpation ($\mu 2$) is:

$$H_0: \mu_1 \geq \mu_2 \text{ vs } H_1: \mu_1 < \mu_2$$

8.3.3.2 Primary Efficacy Analysis

This hypothesis will be evaluated using a bootstrap analysis with a one-sided 0.025 test to indicate statistical significance. The results will be presented for the difference in means and the associated 95% CI; and the ratio of the VerTouch mean divided by the palpation mean and the associated 95% CI.

The primary analysis will include the total insertions for each subject regardless of success on the final placement attempt.

8.3.3.3 Sensitivity and Supporting Analyses

Several supporting analyses are planned for this study:

- The primary analysis will be performed only for subjects who had a procedure success.
- If there are missing data, best- and worst-case imputation analyses will be performed. For the worst-case imputation, the highest observed value will be imputed in the VerTouch group and the lowest value in the palpation group. The best-case analysis reverses the pattern of imputations.
- A Wilcoxon rank-sum test will be used to evaluate the null hypothesis that median attempts are the same in both arms using a two-sided test with a p-value less than or equal to 0.05 considered significant.
- Homogeneity of the study results will be assessed using an ANOVA with site, treatment, and site-treatment interaction terms. A p-value for the site-treatment interaction of less than or equal to 0.15 will be considered statistically significant for this analysis. The n, mean, and SD of values by study site will be provided by treatment group.
- The observed attempts will be evaluated through a t-test for unequal variances comparing the difference in the log (insertion attempts) for each group. The difference will be transformed with the exponent function to provide the ratio of the geometric mean attempts and a two-sided 95% CI for the ratio.

8.3.3.4 Covariate and Other Supporting Analyses

The ANOVA model based on the log(attempts) will consider each of the following covariates one at a time with the factor, treatment, and factor-treatment interaction. The n, mean, and standard deviation (SD) of values for the number of attempts by factor levels will be provided by treatment group. Subgroup levels will only be considered if there are at least 8 subjects within the covariate level across both treatments. The covariates to be analyzed are:

- BMI (kg/m2): underweight (<18.5), normal (18.5-24.9), overweight (25-29.9), class 1 obesity (30.0-34.9), class 2-3 obesity (≥35 .0)
- Age (year): adolescent (<22), young adult (22-40), middle-aged adult (41-65), older adult (>65)
- Provider type: MD, PA, CRNA, AA, NP
- Provider specialty: emergency medicine, neurology, anesthesiology
- MD level (MDs only): resident, fellow, attending
- Provider years in practice: <5, 5-10, >10
- Provider experience (number of neuraxial procedures in past 12 months): 5-25, 26-50, 51-100, >100
- Neuraxial procedure: diagnostic LP, therapeutic LP, neuraxial anesthesia, blood patch
- Medication given for procedure: none, pain, anti-anxiety
- Needle type: spinal 20-22G cutting, spinal 20-22G non-cutting, spinal >22G, Tuohy, CSE
- VerTouch workflow (VerTouch subjects only): marking, introducer placement, needle placement

8.3.4 Secondary and Tertiary Efficacy Endpoints

If that analysis reaches statistical significance, the hypotheses of the secondary variables found in section 8.2 will be considered sequentially such that if one analysis is statistically significant, then the next hypothesis can be evaluated to control for the Type I. The endpoints for these analyses are:

- Incidence of first-insertion success (a case that does not require any reinsertions, but can include any number of redirections)
- Number of redirections (any forward movement of the needle in a new direction not preceded by withdrawal from the skin, counted until confirmation of spinal canal access can be assessed)
- Number of passes (any forward movement of the needle, calculated as the sum of insertions and redirections)
- Incidence of first-pass success (a case that does not require reinsertions or redirections)
- Subject discomfort during landmarking on a 0-10 numeric rating scale (NRS)
- Provider confidence with the identified insertion site on a 1-5 Likert scale
- Procedure success as confirmed by the following procedure-specific methods:
    - Epidural anesthesia: able to achieve a T10 or greater bilateral sensory level change to cold
    - Spinal anesthesia: able to achieve sensory blockade to surgical stimulus at level desired
    - Diagnostic and therapeutic LP: return of CSF
    - Blood patch: able to inject homologous blood into epidural space (entry confirmed by loss-of-resistance)

The analysis methods are outlined Section 8.2 above.

8.3.5 Other Statistical Analyses

There will also be several tertiary and exploratory endpoints examined in this study. The tertiary endpoints are as follows:

- Localization time (time from first touch of draped patient to identification of an insertion site; for VerTouch, this is the time from VerTouch device placement to movement of the applicator to the identified insertion site)
- Insertion time (time from retrieval of marker or LA assembly until no further needle advancements are made)
- Number of bone contacts (counted until confirmation of spinal canal access can be assessed)
- Incidence of referral to radiology

Exploratory endpoints:

- Procedure time (from positioning of the patient to removal of the drape from the subject's back)
- Incidence of conversion from spinal to epidural (specific to neuraxial anesthesia)

The tertiary and exploratory endpoints will be summarized descriptively. Tertiary endpoints of localization time, insertion time, and number of bone contacts will also be evaluated for homogeneity of the group results using a Mann-Whitney test. The exploratory endpoint of procedure time will also be evaluated with a Mann-Whitney test. The tertiary and exploratory incidence endpoints will be evaluated for homogeneity of the rates across the treatment groups using likelihood ratio tests.

8.3.6 Safety Endpoints

The following safety related endpoints were identified in the protocol:

- Incidence of post-dural puncture headache (PDPH)
- Incidence of unintended dural puncture (specific to epidural anesthesia)
- Incidence of paresthesia during needle insertion (specific to neuraxial anesthesia)
- Incidence of traumatic tap (results in visible blood aspiration)

These would be analyzed as binary endpoints and compared with likelihood ratio chi-squared tests.

8.3.7 Adverse Events and SAEs

Adverse events will be summarized using the number and percentage of subjects with one or more events and well as the total count of events. Adverse events will be summarized overall, by device relatedness, seriousness, and severity. Adverse events will only be summarized descriptively.

9 References

1. Chin, KJ, Perlas, A, Chan, V, Brown-Shreves, D, Koshkin, A, & Vaishnav, V (2011). Ultrasound imaging facilitates spinal anesthesia in adults with difficult surface anatomic landmarks. *Anesthesiology*, 115(1), 94-101.
2. Cho, YC, Koo, DH, Oh, SK, Jeong, WJ, Lee, WS, You, YU, & Yoo, IS (2009). Comparison of ultrasound-assisted lumbar punctures with lumbar puncture using palpation of landmarks in aged patients in an emergency center. *Journal of Korean Society of Emergency Medicine*, 20(3), 304-309.
3. Grau, T, Leipold, RW, Conradi, R, Martin, E, & Motsch, J (2002). Efficacy of ultrasound imaging in obstetric epidural anesthesia. *Journal of Clinical Anesthesia*, 14(1), 169-175.
4. Lee, WS, Jeong, WJ, Yi, HY, Ryu, S, Lee, JW, & Kim, SW (2008). The usefulness of ultrasound-assisted lumbar puncture on adult patients in the emergency center: comparison with classic lumbar puncture. *Journal of Korean Society of Emergency Medicine*, 19(1), 562.
5. Mofidi, M, Mohammadi, M, Saidi, H, Kianmehr, N, Ghasemi, A, Hafezimoghadam, P, & Rezai, M (2013). Ultrasound guided lumbar puncture in emergency department: time saving and less complications. *Journal of Research in Medical Sciences*, 18(4), 303-307.
6. Nomura, JT, Leech, SJ, Shenbagamurthi, S, Sierzenski, PR, O'Connor, RE, Bollinger, M, & Gukhool, JA (2007). *Journal of Ultrasound Medicine*, 26(1), 1341-1348.
7. Peterson, MA, Pisupati, D, Heyming, TW, Abele, JA, & Lewis, RJ (2014). Ultrasound for routine lumbar puncture. *Academic Emergency Medicine*, 21(1), 130-136.
8. Singla, P, Dixon, AJ, Sheeran, JL, Scalzo, D, Mauldin, FW & Tiouririne, M (2019). Feasibility of spinal anesthesia placement using automated interpretation of lumbar ultrasound images: a prospective randomized controlled trial. *Journal of Anesthesia & Clinical Research*, 10(2).
9. Wan, X., Wang, W., Liu, J, & Tong, T (2014). Estimating the sample mean and standard deviation from the sample size, median, range, and/or interquartile range. *BMC Medical Research Methodology*, 14(135). Retrieved from: https://bmcmedresmethodol.biomedcentral.com/track/pdf/10.1186/1471-2288-14-135

10 List of Tables, Figures and Listings

Shell Number	Title	Population
14.1.1	Etc.	

To be completed with the mock outputs.

11 Table, Figure, and Listing Shells

[Usually provided in a separate file]