

A Randomized Trial of an Advanced Pneumatic Compression
Device vs. Usual Care for Head and Neck Lymphedema

Protocol Number: 8030
NCT Number: NCT04797390

Statistical Analysis Plan

August 23, 2023
Version 1.1

Statistical Analysis Plan

1 Table of Contents

1	Table of Contents	1
2	Abbreviations and Definitions	2
3	Introduction	2
3.1	Preface	2
3.2	Scope of the analyses	2
4	Study Objectives and Endpoints	2
4.1	Study Objectives	3
4.2	Endpoints	3
5	Study Methods	4
5.1	General Study Design and Plan	4
5.2	Inclusion-Exclusion Criteria and General Study Population	4
5.3	Randomization and Blinding	5
5.4	Study Assessments	5
6	Sample Size	6
7	General Analysis Considerations	7
7.1	Timing of Analyses	7
7.2	Analysis Populations	8
7.2.1	Full Analysis Population (or Intention to Treat)	8
7.2.2	Per Protocol Population	8
7.3	Covariates and Subgroups	8
7.3.1	Multi-center Studies	8
7.4	Missing Data	9
7.5	Multiple Testing	9
8	Summary of Study Data	9
8.1	Demographic and Baseline Variables	9
8.2	Subject Disposition	10
8.3	Derived variables	10
8.4	Protocol Deviations	10
9	Primary Analyses	10
9.1	Secondary Analyses	11
9.2	Exploratory Efficacy Analyses	11
10	Reporting Conventions	11

2 Abbreviations and Definitions

Provide a list of the abbreviations and acronyms used in the Statistical Analysis Plan (SAP) with definitions. All terms will appear in alphabetical order.

This section should be completed on an ongoing basis during the preparation of the document and checked carefully after preparing the rest of the SAP to ensure that all the abbreviations are captured. MS Word has a Find tool that can search for wild cards; <[A-Z]{2,}> will pick up acronyms.

Although the abbreviations are listed, it is standard practice to spell out abbreviated terms and to indicate the abbreviation in parentheses at their first appearance in the text.

AE	Adverse Event
APCD	Advanced Pneumatic Compression Device
CRF	Case Report Form
CT-LEFAT	Computed Tomography Lymphedema and Fibrosis Assessment Tool
HNC	Head Neck Cancer
LEF	Lymphedema and Fibrosis
IMP	Investigational Medical Product
PRO	Patient Reported Outcome
SAP	Statistical Analysis Plan

3 Introduction

3.1 Preface

Lymphedema is one of the common complications associated with head and neck cancer and its treatment. In addition to the detrimental effect lymphedema can have on patients, if left unmanaged it contributes to the development of fibrotic tissue that limits motion and can be extremely detrimental to patients' physical functioning and quality of life. This study is a randomized trial of an APCD versus usual care involving referral to a lymphedema therapist for the treatment of lymphedema among survivors of head and neck cancer with the primary goal of establishing whether there is a discernable benefit to treatment with APCD.

3.2 Scope of the analyses

These analyses will include data collected for subjects enrolled in the Randomized Trial of an Advanced Pneumatic Compression Device vs. Usual Care for Head and Neck Lymphedema Study.

4 Study Objectives and Endpoints

4.1 Study Objectives

To compare the effectiveness of an APCD to Usual Care in the management of LEF in HNC survivors.

4.2 Endpoints

To compare the short-term (2 months) and long-term (4 & 6 months) effectiveness of self-administered APCD therapy to Usual Care in HNC survivors with treatment naïve LEF.

Anatomical Measures:

1. Internal lymphedema
 - Assessment of upper aerodigestive tract (UADT) via endoscopy with direct or indirect visualization (Modified Patterson Scale)
 - Computed Tomography (CT) Imaging (CT Lymphedema and Fibrosis Assessment Tool [CT-LEFAT])
2. External lymphedema
 - Assessment using Head and Neck Cancer Related Lymphedema and Fibrosis Grading (HN-LEFG) criteria
 - Digital Photography (30 section grid)

Patient Reported Biopsychosocial Measures:

1. Symptom Burden: Lymphedema Symptom Intensity and Distress Survey – Head and Neck (LSIDS-H&N v2.0)
2. Symptom Burden and Functional Impairment: Vanderbilt Head and Neck Symptom Survey plus General Symptom Survey (VHNSS plus GSS)
3. QOL: 5-item Linear Analog Self-Assessment (5LAS)
4. Work and Activity: Work Productivity and Activity Impairment Questionnaire (WPAIQ)
5. Perceived self-management compacity: Perceived Medical Condition Self-Management Scale (PMCSMS)
6. Body image: Body Image Quality of Life Inventory (BIQLI)
7. Diet modifications: Automated Self-Administered 24-Hour Dietary Assessment Tool (ASA24*)

5 Study Methods

5.1 General Study Design and Plan

This study is a parallel arm, open label randomized trial of APCD versus usual care. The study has been designed to confirm findings of a pilot trial in which the APCD showed superiority to usual care on a number of PRO endpoints. The overall objective of the study is to conduct a randomized clinical trial comparing the effectiveness of APCD and Usual Care for the management of LEF. The long-term goal is to improve outcomes for individuals suffering from LEF.

5.2 Inclusion-Exclusion Criteria and General Study Population

Inclusion Criteria

1. Age ≥ 18 years
2. Pathologically confirmed cancer of the head and neck (larynx, pharynx, oral cavity, paranasal sinuses, major salivary glands, and HNC of unknown primary)
3. Completed curative intent cancer therapy with no evidence of active cancer at time of study enrollment
4. A diagnosis of either internal or external head and neck lymphedema
5. At least one core lymphedema associated symptom of ≥ 4 out of 10 at the time of study screening
6. Must be able and willing to participate in all aspects of the study and provide informed consent prior to study participation
7. Must be able to speak and understand English

Exclusion Criteria

1. Previous APCD or Usual Care treatment for HNC associated LEF
2. Acute facial infection (e.g., facial or parotid gland abscess)
3. Known carotid sinus hypersensitivity syndrome
4. Symptomatic carotid artery disease, as manifested by a recent (within 30 days of informed consent) transient ischemic attack, ischemic stroke, or amaurosis fugax (monocular visual ischemic symptoms or blindness)
5. Internal jugular venous thrombosis (within 3 months of informed consent)
6. Patient is pregnant or trying to become pregnant.

5.3 Randomization and Blinding

Randomization codes will be generated by the study statistician in a permuted block design. The block size will be balanced within each block and will maintain a 1:1 ratio between treatment groups. Randomization within each site will proceed in 10-unit blocks with a randomly generated sequence defined a priori. This trial is open label by necessity.

5.4 Study Assessments

Aim 1: Anatomic Measures of LEF

Internal Lymphedema:

Assessment of UAD with Modified Patterson Scale: Scoping Procedure. The Patterson Scale was developed to assess edema/swelling in the pharynx and larynx for patients treated with radiation. The scale includes 11 structures and 2 spaces. The scale has good intrarater reliability (weighted kappa, 0.84) and moderate interrater reliability (weighted kappa, 0.54). Four grades are used to evaluate the internal edema level, in the pharynx and larynx, from normal (no edema) to severe edema. We have previously reported findings that internal lymphedema captured by the Patterson Scale correlated with patient reported swallowing difficulty as well as objective findings of swallowing dysfunction on modified barium swallow.

CT Imaging. CT images are used as part of standard of care to evaluate response to oral cavity and oropharyngeal cancer treatment in the study sites. A standardized protocol will be used to score LEF status post head and neck cancer treatment. This will be done centrally, and the reviewer(s) will be blinded to site and randomization group. Imaging will be considered timely if it is done within a 30-day window prior to/post study visit. Our completed R01 study reveals that approximately 70% of the patient's CT scans were done within this window.

External Lymphedema:

Head and Neck Cancer Related Lymphedema and Fibrosis Grading Criteria (HN-LEFG).

Trained study personnel will complete a physical examination of the head and neck area. Grading will be documented via the Head and Neck Lymphedema and Fibrosis Assessment criteria

Digital Photography: Using well-established procedures that have been successful in both single and multi-site studies, research team members will photograph patient's right and left profiles and front from the shoulders to top of head. Patients will be seated against a plain background and be asked to wear a tank top. Photos will be uploaded to a centralized portal to facilitate centralized scoring.

Aim 2: Patient Reported Outcome Measures:

Lymphedema Symptom Intensity and Distress Survey v 2 (L SIDS-Head and Neck v2): LSIDS-H&N). This 31 item tool captures LEF-related symptom burden in HNC patients. It includes 7 symptom clusters. Internal consistency values of scores generated from the HNC symptom clusters range from 0.83 to 0.95 confirming the cluster compositions. The LSIDS-H&N V2.0 (manuscript in-press) takes 5-7 minutes to complete.

Vanderbilt Head and Neck Symptom Survey plus General Symptom Survey version 2.0 (VHNSS v2.0 plus GSS): The VHNSS v.2.0 consists of 50-items within 13 domains including nutrition, swallowing, xerostomia, mucositis, excess mucus, speech, hearing, taste change, smell, dental health, mucosal sensitivity, range of motion, and pain. Items are scored on a numeric scale rating the severity of the symptom from 0 (none) to 10 (severe). VHNSS v2.0 plus GSS includes 12 additional items directed at the systemic effects of therapy. Studies have reported good internal consistency for the total scale ($\alpha=0.94$) and five subscales (i.e., swallow, nutrition, mucous/dry mouth, pain, voice) (α s=0.77-0.93), and adequate convergent and divergent validity. The VHNSS v2.0 plus GSS takes approximately 10 minutes to complete.

Quality of Life: The 5-item Linear Analog Self-Assessment (5LAS): This assessment will be used to evaluate quality of life. The scale assesses physical, emotional, spiritual, intellectual, and overall well-being, on a 0-10 scale. It has been used successfully in cancer populations. It takes approximately <1 minute to complete

Work Productivity and Activity Impairment Questionnaire (WPAIQ): This commonly used, valid and reliable, 9-item questionnaire evaluates lost hours of productivity seven days prior to administration and can be targeted to evaluate lost productivity due a specific medical condition (e.g., lymphedema). When compared to the SF-36, it has been found to correlate with general health scores ($r=0.52$), and physical role ($r=0.52$).

Perceived Medical Condition Self-Management Scale (PMCSMS): The PMCSMS is a generic (template) instrument that can be specific to any medical condition. The scale was initially used in individuals with diabetes ($N=398$) and Cronbach's α was .84 and has since been validated against a shorter version of the instrument that did not perform as well.

Body Image Quality of Life Inventory (BIQLI): This validated, self-report measure captures body image (19 items, Cronbach's $\alpha = .95$).

Automated Self-Administered 24-Hour (ASA24®) Dietary Assessment Tool (REF): Detailed information about dietary intake and nutritional values can be obtained using this NCI tool. It can be completed using smart phones, tablets, laptops, and desktops. Patients will enter their information and the program will calculate nutrition results. LEF has been associated with dysphagia and dietary adaptations. This will allow us to determine whether improvement in swallow function can result in improvement in dietary intake.

6 Sample Size

Preliminary data from a previous study, which was administered in patients whose LEF was refractory to usual care, to perform the power calculations for this study. Because the previous study selected for patients who had already failed usual care, at least for a time, we expect that the effect sizes in that population were somewhat attenuated compared to what we would see in this study's treatment naïve population. Thus, we expect that the power calculated here is slightly conservative.



Similar parametric calculations will verify that at a sample size of approximately 210, the study has >80% power to detect a 1-unit change on a symptom subscale which has an SD of 3. Changes of 1 unit are generally clinically meaningful and SDs of approximately 3 units are common among VHNSS items in this population as demonstrated in the table above. This mirrors the upstrap calculations which gave a similar size of 200 with the balance of the planned patients justified to account for dropouts.

7 General Analysis Considerations

7.1 Timing of Analyses

Once the required sample size has been followed to a study endpoint (2 months, 6 months) and all data validation and cleaning procedures have been performed to that point, the data set will be locked and subsequently analyzed. Analysis of the control group only to

determine the proportion of patients who received lymphedema therapy and its timing may occur without breaking blinding as no primary outcome data is required to conduct it and therefore may take place at any time including prior to data lock.

7.2 Analysis Populations

7.2.1 Full Analysis Population (or Intention to Treat)

- All subjects who were randomized and have data at baseline and follow-up data inclusive of the endpoint being analyzed.

7.2.2 Per Protocol Population

- All subjects who did not substantially deviate from the protocol as to be determined on a per-subject basis at the principal investigators' discretion at the time of database lock. Examples of substantial deviations would be arm switching, adoption of a disqualifying treatment, or similar.

7.3 Covariates and Subgroups

It is anticipated that patients in both arms of the study will experience some variability in the efficacy of their assigned therapy according to fidelity. However, given the increased functional and logistical burden associated with scheduling and attending in-person sessions with a lymphedema therapist, it is anticipated that there will be substantially more variability in the usual care arm. In addition, disease specific information such as the location, cancer stage, treatment administered as part of cancer therapy, and baseline LEF characteristics may have bearing on the state of LEF at enrollment as well as its response to therapy. We expect there to be correlation between outcomes by site. Potential differences by sex and race will be investigated but are not necessarily anticipated. These demographic and disease related covariates will be balanced between groups with a high probability given the study's size and design, however, balance in these areas will be investigated and considered exploratory...

Effect modification by demographic and disease specific information will be investigated by introducing an interaction term into the model for the primary analysis. Statistical significance of the interaction term, accompanied by a coefficient of clinically meaningful size, will be taken as evidence supporting a differential treatment effect. These investigations will be considered exploratory.

7.3.1 Multi-center Studies

It is not anticipated that any center data will need to be combined with data from other centers. However, if a center contributes less than ten patients, this will be considered to the extent that it improves the mathematical properties of the statistical model proposed for the primary analysis. If this condition is met, results for each center will be compared with the results from the under-represented center, and its data will be combined with the center's which is most similar and is likely to produce the most appropriate mathematical qualities.

The potential for differential effects by study site will be investigated, if evidence of such difference exists. Because the primary analysis will use a mixed-effects model with a random effect for site, it will also produce an estimate of the variance by site and a statistical significance of that variance being non-zero. If there is substantial evidence of non-zero variance by site, and the estimate of the variance by site is of such a magnitude that degree of observed heterogeneity represents an effect size that is likely to be clinically meaningful, as determined by the PI, we will investigate site specific factors which may have contributed to the heterogeneity, including fidelity to the device, successful initiation of therapy with a lymphedema therapist, fidelity to lymphedema therapy, and disease related covariates.

7.4 Missing Data

Missingness will be assessed in all data fields used in the analysis and described both numerically and graphically. Unless evidence to the contrary exists, data will be assumed to be missing at random. Missing outcome data will not be imputed. Covariates that have missing values will be imputed. If less than 10% of the values for a covariate are missing, then single conditional mean imputation may be considered. For analyses that require a covariate that is missing in $\geq 10\%$ of participants, multiple imputation will be employed along with Rubin's rules for variance estimation.

Dropout rate and timing will be assessed by treatment arm to assess the potential for differential dropout, which might impact the study results.

7.5 Multiple Testing

Despite the wide range of potential outcome variables, no multiple testing adjustment is planned due to the unique rationale for the study. Multiple comparison adjustments are primarily done to protect an alpha-level type 1 error rate in a decision to declare superiority of the treatment group over the control group. Because this study is being conducted, not as a registration study, but in an effort to demonstrate that the intervention has the capacity to improve clinical signs and symptoms across a wide range of variables, effort, there is no need for a declaration of superiority compared to the control group. Because the definition of what constitutes a 'successful trial' in this case is nebulous with positive decisions being possible even if the treatment and control groups prove to be comparable over a broad selection of outcomes, it is not possible to protect the error rate of this decision. Thus, the best alternative is to present the data for each independent outcome as it stands with independent $\alpha=0.05$ error rates.

8 Summary of Study Data

8.1 Demographic and Baseline Variables

Demographic variables will include, at a minimum, age, sex, race, ethnicity, BMI, and study location. Pertinent baseline variables include measurements from imaging, LSIDS scores,

VHNS scores, QOL scores, HNLFEQ, WPAIQ scores, and cancer disease and treatment variables.

This data will be stratified according to the ITT control and intervention groups. Summaries of these variables will be given as mean (sd) for continuous variables and n (%) for categorical variables. Statistical measures are generally not advisable because they often prove misleading to the reader, but are often insisted upon by journals. If such comparisons prove necessary for publication, continuous variables will be analyzed using a standard t-test, and categorical variables will be analyzed using a chi-squared or Fisher's Exact test as appropriate.

8.2 Subject Disposition

We will follow CONSORT guidelines in the reporting of subject disposition. A traditional consort diagram containing levels for screening, randomization, allocation/baseline visit, each of the 3 follow-up times and analysis will be generated describing the number of participants in each arm at each level. Loss to follow-up will be deemed to occur at the level of the first missing visit which is not followed by a subsequent visit.

8.3 Derived variables

Items on each patient reported outcome measure will be analyzed according to the tool's authors recommendation, in cases where such a recommendation exists. For the VHNS and the LSIDS, this will involve averaging the response items in a given subscale. However, there have been cases where alternative scoring has been used (i.e., max item score on a subscale) for the VHNS and any analyses that use such a transformation should be considered exploratory.

8.4 Protocol Deviations

Deviations from the assigned treatment arm have been discussed in the section on ITT versus per protocol analysis. In summary, the primary analysis of all study outcomes will be done as an ITT analysis. Any analyses conducted on the per protocol population will be considered exploratory. All protocol deviations including missed visits, out of window visits, and failure to receive lymphedema therapy will be analyzed descriptively.

9 Primary Analyses

The primary endpoints for this study are broken into two categories, clinical measures and patient reported measures. These endpoints will each be measured in a similar fashion using a mixed-effects generalized linear regression model with a link function chosen to match the outcome type. For continuous measures including CT measurements, PRO overall scores, PRO subscale scores, and particular PRO items, the identity link will be used. The design of this study suggests correlation by patient and by study site. These factors will have random intercept terms included in the model to account for this correlation. For each outcome, the appropriate mixed-effects regression model will be fit with a term for treatment assignment, baseline score, and the aforementioned random-effects.

A significant coefficient associated with the treatment assignment variable would indicate evidence of a difference between the treatment groups, [REDACTED]

██████████ The mean change from baseline will be reported along with its 95% confidence interval. ██████████ changes in PRO from baseline will be reported graphically in forest-type plots for easy visualization of the results and comparison to the changes observed in the treatment group.

It is not anticipated that further model conditioning will be necessary in this study given its relatively large size. However, if meaningful differences in relevant covariates are suspected after the aforementioned analysis of demographic and baseline variables, other covariates may be included to minimize any potential for confounding which avoided randomization.

9.1 Secondary Analyses

The primary analysis will be repeated with a full set of covariates to include age, sex, BMI, baseline LEF measures, and cancer treatment related variables to assess any difference between the crude and adjusted models. These analyses will similarly consist of generalized linear mixed-effects regression models with random intercepts for participant and study site. Independent analyses of 2-month and 6-month outcomes will also be conducted..

9.2 Exploratory Efficacy Analyses

All analyses of outcomes not explicitly listed in the protocol as primary will be considered exploratory. This includes alternative transformations of primary outcome variables (such as max (subscale score)) and analyses conducted per protocol. Time to receipt of lymphedema therapy amongst those in the control arm will also be analyzed using a Kaplan-Meier approach.

In addition, sensitivity analyses may be considered if they are deemed beneficial to the presentation of the study. Examples may include: tipping point analyses in the case of extreme missingness or suspect levels of differential dropout and alternative modeling approaches (such as ordinal models for the analysis of PRO data in cases where the modelling assumptions are suspect).

10 Reporting Conventions

P-values ≥ 0.001 will be reported to 3 decimal places; p-values less than 0.001 will be reported as “<0.001”. The mean, standard deviation, and any other statistics other than quantiles, will be reported to one decimal place greater than the original data. Quantiles, such as median, or minimum and maximum will use the same number of decimal places as the original data. Estimated parameters, not on the same scale as raw observations (e.g., regression coefficients) will be reported to 3 significant figures.