

**Statistical Analytic Plan for Symptom Screening Linked to Care Pathways
for Children with Cancer: A Cluster Randomized Trial**

Study Title: Symptom Screening Linked to Care Pathways for Children with Cancer: A Cluster Randomized Trial

NCT Number: NCT04614662

Document Date: November 21, 2023

Objectives	2
Sample Size Justification	2
Analytic datasets	2
Study Outcomes	3
Descriptive Statistics	5
Comparative Analyses	5
Primary outcome	5
Secondary outcomes	5
Missing data	7
Software	9

Objectives

Aims 1 and 2: Among children with newly diagnosed cancer, to determine if symptom screening and feedback to healthcare providers at least three times weekly and locally-adapted symptom management care pathways, when compared to usual care:

Aim 1. Improves overall self-reported symptom scores (total SSPedi score), fatigue (PROMIS–Fatigue) and cancer-specific quality of life (QoL) (PedsQL 3.0 Acute Cancer Module) over 8 weeks.

Hypothesis: Symptom screening and care pathways will improve symptoms, fatigue and QoL

Aim 2. Improves symptom documentation, increases provision of interventions for symptoms, and reduces emergency department visits and unplanned clinic visits and hospitalizations over 8 weeks.

Hypotheses: Symptom screening and care pathways will increase symptom documentation and provision of interventions for symptoms, and will reduce healthcare utilization.

Aim 3: As an exploratory aim, we will evaluate key elements of the intervention related to the external validity and generalizability of the intervention effects using the RE-AIM framework.

Sample Size Justification

The primary endpoint is the total self-reported SSPedi score at eight weeks. We used data from our SSPedi validation study for power calculation. In that study, we recruited 302 children across nine sites who were hospitalized and expected to be in hospital or in clinic three days later. Children completed SSPedi on days 1 and 4 and a 5-point global symptom change scale on day 4. The average absolute change in total SSPedi scores for those who reported they were a little better or a little worse on this global change scale (a 1-point change) was 2.7 while the average absolute change in total SSPedi scores for those who reported they were much better or much worse (a 2-point change) was 5.6. Thus, the minimum plausible clinically important difference is likely approximately 3.0 but may be higher. The intracluster correlation coefficient (ICC), a measure of the degree of clustering, was 0.021 (95% confidence interval (CI) 0 to 0.123). In power calculations for the proposed study, we used ICC=0.021 but show sensitivity analyses to ICCs at the extremes of the 95% CI (which are unlikely). Among the 20 sites in the proposed trial, we anticipate that four will each contribute eight patients, four will each contribute 32 patients, and the remaining 12 will each contribute 20 patients, for a total of 400 patients. With these cluster sizes, Table 1 shows power for combinations of these ICCs and clinically important differences, at a 2-sided alpha of 0.05 and assuming a within-cluster standard deviation of 8.8 (derived from the SSPedi validation study). Power for each combination in the table was calculated through simulation of 4000 trials. Each one was analyzed with a linear mixed effects model clustering by site and the percentage with a significant P value was taken as the estimate of power. If there is a baseline characteristic, such as age, that explains 20% of the variance in 8-week scores, including that variable in analyses increases power to the value shown in the far right column. Table 1 shows that assuming a minimal clinically important difference of 3.0 and with inclusion of age as a covariate in the analysis, there is 85% power with ICC=0.021. Assuming that 10% of patients have missing final

scores, we inflated the sample size to 444.

Table 1: Power With 400 Patients, Alpha=0.05 and Standard Deviation=8.8			
ICC	Difference	Power (%)	
		Without baseline predictor	With age as baseline predictor
0.000	2.5	71.0	81.4
	3.0	86.7	94.2
	3.5	95.9	98.9
0.021	2.5	59.5	70.1
	3.0	75.7	85.2
	3.5	86.5	94.2
0.123	2.5	26.6	32.6
	3.0	38.2	45.3
	3.5	47.7	57.0

Analytic Datasets

All analyses will follow the intention-to-treat principle. See below (Missing Data) for details on how those with missing data will be included in the analysis.

Study Outcomes

Questionnaires including the primary outcome and secondary patient-reported outcomes are obtained at weeks 4 and 8 from enrollment.

1. Primary outcome: The primary outcome is the total SSPedi symptom score, which is the sum of each of the 15 SSPedi item's Likert scores, resulting in a total score that ranges from 0 (no bothersome symptoms) to 60 (worst bothersome symptoms). The designated week 4 or 8 SSPedi score will be taken on the day the other patient-reported outcomes were obtained or were planned to be obtained. This assures that the SSPedi score was collected as an outcome rather than as an intervention.
2. Secondary outcomes:
 - a. Individual SSPedi symptom scores, which consist of each of the 15 symptoms scored as 0, 1, 2, 3 or 4.
 - b. Fatigue will be measured using PROMIS. The raw score is translated to a T-score for each participant, where in the general population, the average T-score is 50 and the standard deviation is 10. The recall period is the last seven days. A higher T-score represents more fatigue, or worse health.
 - c. PedsQL 3.0 Acute Cancer Module (scored according to <https://www.pedsq.org/PedsQL-Scoring.pdf>), which produces 8 domain scores on a 0-100 scale. The domains are (1) pain and hurt; (2) nausea; (3) procedural anxiety; (4) treatment anxiety; (5) worry; (6) cognitive problems; (7) perceived physical appearance; and (8) communication. The recall period is the last seven days. Higher scores indicate better QoL.

- d. Documentation of symptoms (mapped onto the 15 SSPedi symptoms) in the participant's health record; these are a set of 15 binary outcomes (assessed at weeks 4 and 8). For each, the window is a 3-day period extending from one day before to one day after the week 4 and 8 SSPedi time point.
 - e. SSPedi symptom-specific intervention performed; these are a set of 15 binary outcomes (assessed at weeks 4 and 8). For each, the window is a 3-day period extending from one day before to one day after the week 4 and 8 SSPedi time point.
 - f. Healthcare encounters including emergency department visits, unplanned clinic visits and unplanned hospitalizations over the 8-week period. The number of encounters in each category and the total across categories will each be a separate secondary outcome.
3. Exploratory outcomes: These are RE-AIM characteristics to measure reach, effectiveness, adoption and implementation.
- a. For reach, we propose to describe the baseline demographic characteristics of eligible patients who participate and who do not participate. We will also describe site attributes using the inner setting measures from the Consolidated Framework for Implementation Research (CFIR) as measured at baseline.(18, 19) This conceptual framework includes factors that may impact on intervention implementation. These attributes include culture, culture stress, culture effort, implementation climate, learning climate, leadership engagement and available resources.
 - b. Effectiveness will be evaluated through primary and secondary outcomes as described in list items 1 and 2 above and also by comparisons of the percentages of patients with severely bothersome symptoms within intervention and control sites. We will characterize adoption at intervention sites by describing characteristics of sites and providers at those sites. Site characteristics will include pediatric vs. mixed adult and pediatric, number of pediatric cancer patients diagnosed annually, and percentage of patients anticipated to have private, public or no insurance. Healthcare professional characteristics will include number of physician and nurse practitioner full-time equivalents and median years in practice.
 - c. Adoption will be assessed through the number of SSPedis per participant completed at intervention institutions; this analysis will use duration of time on protocol therapy as an offset. Further, we will dichotomize each participant as a high or low SSPedi completer, with a high completer being defined as one who completed 15 or more SSPedis. This analysis will use those who complete protocol therapy as planned as the denominator. In addition, we will count the number of times the institutional care pathways were clicked at intervention sites. We will describe the total number of clicks overall and by specific care pathway, access route (email link or QR code) and device (iPad, iPhone, Mac, Windows or unknown).
 - d. Implementation outcomes will capture the percentages of intervention patients who came off protocol therapy early and the number of all participants who

came off study early, both overall and stratified by intervention or control group.

Descriptive Statistics

Baseline characteristics: Summary statistics on baseline characteristics will be calculated within each study arm and presented in a table – means and standard deviations (SDs) for variables treated as continuous, counts and percentages for categorical variables, along with total numbers with missing data.

Outcomes: Similar tables will be created for all study outcomes. Individual SSPedi symptom scores will be summarized by the percentages of participants with scores of 3 or 4, the two most bothersome categories.

Comparative Analyses

Primary Outcome

The primary analysis will use the patient-level 8-week total SSPedi score as the outcome in a mixed linear regression model with a random effect for cluster, fixed effects for treatment assignment, child age and diagnosis group (leukemia/lymphoma, solid tumor and central nervous system tumor), and two binary fixed effects at the cluster level for stratification variables (above or below the average anticipated yearly number of English or Spanish-speaking cancer patients 8-18 years of age, and above or below the average anticipated percentage of patients with private insurance vs. other payment types). Child age and diagnosis group will be included as they are known to be associated with total SSPedi scores; this will reduce residual variance in scores. We do not plan to adjust the model for baseline SSPedi score because some of the care pathway interventions (that are prophylactic) will affect symptoms prior to enrollment and the baseline score may already reflect some benefit of the cluster's treatment assignment. The treatment effect will be the covariate-adjusted mean difference between study groups in 8-week SSPedi, presented with a 95% CI and a two-sided p-value for a test of no treatment effect. If we find the residuals from the mixed model fitted to the 8-week total SSPedi score deviate substantially from normality, we will conduct sensitivity analyses, one using normalizing transformations of the outcome and another using an ordinal regression model.

To account for potential differential enrollment of surgery only patients (who might not be cared for by oncology teams) between intervention vs. control sites, we will also conduct a sensitivity analysis restricted to patients that received systemic chemotherapy.

Secondary Outcomes

1. **Self-reported individual symptoms at weeks 4 and 8 on the SSPedi instrument:** For each of the 15 ordinal outcomes representing symptoms (scored 0-4), a mixed effects proportional odds model will be fitted using treatment assignment and stratification factors as fixed effects covariates and with a random effect for site. As some individual symptoms may be uncommon, to avoid specifying models that may be overfitted for those symptoms, these models will use only this minimal set of three covariates. The estimated odds ratio for the intervention (representing the odds of having a higher vs. lower score in the intervention group) will be estimated and presented along with a 95%

CI. Two figures will be generated to present these results graphically: (a) to display the raw outcome data, a set of 15 pairs of stacked barplots will be created; within each pair, the stacked barplot will show the percentages of participants with scores of 0-4 in the control and intervention groups; (b) a forest-plot type figure will be created showing the estimated odds ratios with 95% CIs for each of the 15 symptoms.

2. **PROMIS Fatigue Scale:** This scale will be treated as a continuous variable and analyzed using a linear mixed effects model in the same way as the primary SSPedi outcome.
3. **PedsQL 3.0 Acute Cancer Module domain scores:** Each of these 8 scores will be treated as continuous variables and analyzed using a linear mixed effects model in the same way as the primary SSPedi outcome. A forest-plot type figure will be created showing the estimated treatment effects with 95% CIs for each of the 8 domains.
4. **Documentation of symptoms on weeks 4 and 8 (± 1 day):** Analyses will involve three analyses for each symptom class. Each analysis will fit a mixed effects logistic regression model to estimate the odds ratio for documentation (on weeks 4 and 8) comparing intervention groups, adjusting for stratification factors. It is possible that there will be too few documentation events to reliably fit this mixed model to all 15 symptoms. If this model cannot be fitted to some symptoms (likely because the random effects variance cannot be estimated), we will fit a fixed effects logistic regression model to just those symptoms. Three cohorts will be defined:
 - a. Each participant will be classed as having the symptom documented or not. All participants are included in the analytic sample for each symptom.
 - b. The analytic sample will be limited to those participants who self-report the symptom on weeks 4 and 8 (symptom score > 0).
 - c. The analytic sample will be limited to those participants who have the most bothersome self-report of the symptom on the weeks 4 and 8 SSPedi (symptom score ≥ 3).
5. **Interventions for symptoms:** Analyses will follow a similar flow to the analyses of documentation of symptoms. For each of the 15 symptoms, each participant will be classified as having an intervention at weeks 4 and 8 (± 1 days). Also collected is whether interventions were administered specifically for the specific symptoms. For example, acetaminophen may be administered for fever or pain. As interventions are expected to be more uncommon than symptoms, all treatment effects will be estimated by simple differences in the proportions with interventions (i.e., ignoring clustering), p-values will be computed from the Fisher exact test, and 95% CI for differences in proportions will use the Newcombe method. Three cohorts will be defined:
 - a. Each participant will be classed as having an intervention for the symptom or not. All participants are included in the analytic sample for each intervention-symptom pair.
 - b. The analytic sample will be limited to those participants who self-report the symptom on the weeks 4 and 8 SSPedi (symptom score > 0).
 - c. The analytic sample will be limited to those participants who have the most bothersome self-report of the symptom on the weeks 4 and 8 SSPedi (symptom score ≥ 3).

6. **Unplanned encounters:** For each of the secondary endpoints of unplanned health services utilization (emergency, clinic visits, hospitalizations, and the total across categories), the number of events per child will be counted over the 8-week period and compared between randomized groups using a mixed-effects Poisson or negative binomial regression analysis (with zero-inflation if necessary), with a random effect for site and fixed effects for the two stratification factors. If a child has less than 8 weeks of follow-up, the model will include an offset equal to the logarithm of each child's follow-up time.

Exploratory Outcomes

1. **Reach:** we will describe the baseline demographic characteristics (sex, age or age group and diagnosis group) of eligible patients who participate and who do not participate using means and standard deviations for continuous variables and counts and percentages for categorical variables. The difference between participants and non-participants will be summarized two ways: numerically by the standardized mean difference and graphically by side-by-side barplots for categorical variables and by side-by-side beeswarm plots for continuous variables. We will also summarize CFIR inner setting attributes as counts and percentages.
2. **Effectiveness:** In addition to the analyses of primary and secondary outcomes detailed above, we will also carry out analyses that dichotomize each of the 15 SSPedi symptoms as being severely bothersome (3 or 4) or not. Percentages for each symptom will be tabulated and compared between intervention groups using mixed effects binary logistic regression. We will also assess between-site variation in the odds of having a severely bothersome symptom (after adjusting for intervention and stratification factors) by plotting the random effects associated with each site and by calculating the median odds ratio (MOR). This is an alternative measure of clustering (or between-cluster variability); it is the estimated median of the ratio of the odds of a bothersome symptom for a participant at one site to the odds of a bothersome symptom for an otherwise identical participant at another site. The MOR is always defined with the higher odds in the numerator. One appealing aspect of the MOR is that it is on the same scale as the odds ratio for treatment, aiding interpretation of both the OR for intervention and the MOR for variation in the outcome.
3. **Adoption:** We will characterize adoption at intervention sites by summarizing the following:
 - a. characteristics of sites and their providers
 - b. number of SSPedis completed per enrolled participant
 - c. proportion of participants that are high or low SSPedi completers
 - d. number of times the institutional care pathways were clicked.

Site and provider attributes will be described by using means and standard deviations for continuous variables and counts and percentages for categorical variables. Site characteristics will include pediatric vs. mixed adult and pediatric, number of pediatric cancer patients diagnosed annually, and percentage of patients anticipated to have private, public or no insurance. Healthcare professional characteristics will

include number of physician and nurse practitioner full-time equivalents and median years in practice. CFIR attributes will also be described.

The dependence of the number of SSPedis completed by a participant on participant, guardian and site characteristics will be assessed using a mixed effects Poisson or negative binomial regression model, with a random effect for site. Participant characteristics will include sex, age/age group, race, ethnicity, preferred language, and cancer diagnosis group. Guardian characteristics will include marital status, guardian employment and highest education, and high annual household income. Site characteristics will be those included above. The dependence of the odds of being a high responders on these same characteristics will be assessed using mixed effects logistic regression adjusting for stratification factors.

The number of recorded clicks is available on aggregate for each site, so we will summarize this as a set of 10 rates, clicks per enrolled participant, and assess whether there is site-to-site variability by fitting a random effects Poisson or negative binomial model with clicks as the outcome and the logarithm of the number of patients as an offset. A summary rate will be estimated from this model and the median rate ratio (analogous to the MOR above) will be calculated.

4. Implementation will calculate the percentage of patients who enrolled but came off protocol therapy early (intervention only) or off study early (all participants), overall and stratified by intervention or control group. Counts and percentages will be shown, with no inferential statistics.

Missing Data

Rule-based methods: The PedsQL will be scored according to its scoring guide, which includes rules for scoring in the presence of missing elements on each domain modules. No other deterministic rule-based method will be used to deal with missing data.

Multiple imputation: If more than 10% of participants are missing the primary outcome or secondary patient-reported outcomes, the analysis will use multiple imputation based on previous SSPedi assessments (weeks 0 and 4), group and site. In addition, we will conduct sensitivity analysis assuming that missing scores are from children with high total SSPedi scores and from children with low total SSPedi scores (randomly drawn from the highest and lowest quartile). A set of 20 complete datasets will be generated using separate within-treatment-group imputation models. These models will include all the study outcomes listed just above, and the key covariates that appear in our adjusted models (age, diagnosis group and stratification factors). The models will be fitted in each imputed dataset and the pooled estimate of the treatment effect and its standard error will be computed using Rubin's rules.

Analyses of documentation and interventions will not use any imputation. There can be no missing data for these outcomes: if a symptom is not documented in the health record and there is no record of an intervention being used, then this constitutes a 'no' for the outcome. While it is possible that the SSPedi symptom score needed for the identification of the analytic datasets in analyses (4b, 4c, 5b, and 5c) is missing for some participants, the more exploratory nature of these analyses led us to the decision to use only observed data for these outcomes.

Software

All analyses will use R 4.3.0 with these specific packages:

- base R for general programming and fitting linear and logistic regression models.
- the tidyverse and lubridate packages for data “wrangling”.
- the rms package for the proportional odds model.
- the mice package for multiple imputation.
- the ggplot2 package for graphs.
- the tableone package for summary descriptive tables.
- glmer for linear and logistic mixed effects regression models
- ordinal for proportional odds mixed effects regression models

Planned Tables

Planned tables are below.

Table 1: Participant Demographic Characteristics

Characteristics	All Patients	Symptom Screening	Usual Care
Male			
Median Age in Years (range)			
Age Group in Years			
8-10			
11-14			
15-18			
Race			
American Indian or Alaska Native			
Asian			
Black or African American			
Native Hawaiian or Other Pacific Islander			
White			
Unknown			
Ethnicity			
Hispanic or Latino			
Not Hispanic or Latino			
Unknown			
Diagnosis			
Leukemia			
Lymphoma			
Solid tumor			
Brain tumor			
Other			
Metastatic Disease			
Median Days from Diagnosis (range)			
Planned or Received Treatment			
Chemotherapy			
Radiotherapy			
Surgery			
Median Days from Treatment Start (range)			
Inpatient at Enrollment			
First language English or Spanish			
Preferred Language for Patient Reported Outcomes			
English			
Spanish			
Not applicable (chart review only)			
Family Composition - Married			
Guardian Employment Full or Part Time			
Guardian Education College or Higher			
Annual Household Income \geq \$60,000			

Table 2: SSPedi Outcomes by Assessment Timepoint and Group

	Baseline		Week 4		Week 8		Treatment Effect Estimate; 95% CI; p-value	
	Symptom Screening	Usual Care	Symptom Screening	Usual Care	Symptom Screening	Usual Care	Unadjusted	Adjusted
Primary Outcome							<u>Mean Difference</u>	
SSPedi Score (mean, SD)								
							<u>Odds Ratio</u>	
Secondary Outcomes								
Severely Bothersome, n (%)*								
Feeling disappointed or sad								
Feeling scared or worried								
Feeling cranky or angry								
Problems with thinking or remembering things								
Changes in how your body or face look								
Feeling tired								
Mouth sores								
Headache								
Hurt or pain (other than headache)								
Tingly or numb hands or feet								
Throwing up or feeling like you may throw up								
Feeling more or less hungry than you usually do								
Changes in taste								
Constipation (hard to poop)								
Diarrhea (watery, runny poop)								

Table 3: Non-SSPedi Patient-reported Outcomes by Assessment Timepoint and Group

	Baseline				Week 8		Treatment Effect Estimate; 95% CI; p-value	
	Symptom Screening (n=175)	Usual Care (n=169)			Symptom Screening (n=168)	Usual Care (n=164)	Unadjusted	Adjusted
PROMIS Fatigue (Mean, SD)							<u>Mean Difference</u>	
Total PedsQL 3.0 Acute Cancer Module Score (Mean, SD)							<u>Mean Difference</u>	
Pain and hurt								
Nausea								
Procedural anxiety								
Treatment anxiety								
Worry								
Cognitive problems								
Perceived physical appearance								
Communication								

Table 4: Symptom Documentation and Intervention by Group for All Participants

	Symptom Screening	Usual Care	P value
Symptom Documentation (%)			
Feeling disappointed or sad			
Feeling scared or worried			
Feeling cranky or angry			
Problems with thinking or remembering things			
Changes in how your body or face look			
Feeling tired			
Mouth sores			
Headache			
Hurt or pain (other than headache)			
Tingly or numb hands or feet			
Throwing up or feeling like you may throw up			
Feeling more or less hungry than you usually do			
Changes in taste			
Constipation (hard to poop)			
Diarrhea (watery, runny poop)			
Any Symptom Intervention (%)			
Feeling disappointed or sad			
Feeling scared or worried			
Feeling cranky or angry			
Problems with thinking or remembering things			
Changes in how your body or face look			
Feeling tired			
Mouth sores			
Headache			
Hurt or pain (other than headache)			
Tingly or numb hands or feet			
Throwing up or feeling like you may throw up			
Feeling more or less hungry than you usually do			
Changes in taste			
Constipation (hard to poop)			
Diarrhea (watery, runny poop)			
Symptom Intervention Clearly			

for Symptom (%)			
Feeling disappointed or sad			
Feeling scared or worried			
Feeling cranky or angry			
Problems with thinking or remembering things			
Changes in how your body or face look			
Feeling tired			
Mouth sores			
Headache			
Hurt or pain (other than headache)			
Tingly or numb hands or feet			
Throwing up or feeling like you may throw up			
Feeling more or less hungry than you usually do			
Changes in taste			
Constipation (hard to poop)			
Diarrhea (watery, runny poop)			

Table 5: Number of Unplanned Healthcare Encounters by Group

Unplanned Encounter Type	Total	Symptom Screening	Control
		(n=xx)	(n=xx)
All unplanned encounters			
Rate per 100 patient-weeks			
Number of encounters (n, %)			
0			
1			
2			
3			
4			
5			
6 or more			
Emergency department visits			
Rate per 100 patient-weeks			
Number of encounters (n, %)			
0			
1			
2			
3			
4			
5			
6 or more			
Unplanned clinic visits			
Rate per 100 patient-weeks			
Number of encounters (n, %)			
0			
1			
2			
3			
4			
5			
6 or more			
Unplanned hospital admissions			
Rate per 100 patient-weeks			
Number of encounters (n, %)			
0			
1			
2			
3			
4			
5			
6 or more			