STATISTICAL ANALYSIS PLAN

ClinicalTrials.gov Identifier: NCT03249090

September 30, 2022

Statistical Analysis Plan for Protocol Title:

Electronic patient reporting of symptoms during outpatient cancer treatment: A U.S. national randomized controlled trial (the "PRO-TECT" trial)

Protocol Title: ***Electronic patient reporting of symptoms during outpatient cancer treatment: A U.S. national randomized controlled trial (the "PRO-TECT" trial)***

Sponsor Name: ALLIANCE FOUNDATION TRIALS (AFT)

NCORP Committee: Cancer Care Delivery Research

This is the preliminary Statistical Analysis Plan (SAP) for the AFT-39 trial. This document will be updated as necessary throughout the life of the trial.

Version No.:	Valid from (date of implementation):	Replacement of:	Number of pages:
1	April 1, 2021	--	**17**
2	June 1, 2022	Version 1	**20**
3	September 30, 2022	Version 2	**21**

I confirm that I have read & approved the above-referenced document for use in the AFT-39 trial.

Written by:

Alliance	*Dueck, Amylou*
Date *(mm/dd/yyyy)*	09/30/2022
Signature	[redacted]

Reviewed by:

Alliance	*Basch, Ethan*
Date *(mm/dd/yyyy)*	09/30/2022
Signature	[redacted]

Summary of Changes/Updates

Version No.	Version Date	Affected Section(s)	Summary of Revisions Made
2	June 1, 2022	Table of contents	Updated to reflect new section(s) and page numbers
		2.2.1	Clarified that health utilities and QALYs are meaningful outcomes separate from overall survival and quality of life; clarified that analysis of emergency room visits includes hospitalizations; added additional endpoints
		2.6	Clarified power calculations
		3.1.1	Removed QALY abbreviation (previously defined)
		3.1.6	Section added to describe how administrative and healthcare utilization databases will be incorporated into the analysis datasets
		3.2.3	Clarified reason for delayed analysis of overall survival; clarified that analysis of emergency room visits includes hospitalizations; clarified plan for additional administrative database downloads; stated censoring plan to standardize follow-up across patients
		3.4.2	Added cancer diagnosis related covariates to better delineate each patient's disease advancement
		3.5.3	Incorporated cancer diagnosis related covariates to primary survival analysis (approach changed from Kaplan-Meier to Cox regression to accommodate new covariates); clarified health utilities and QALY analyses
		3.5.4	Added sensitivity analysis censoring patients at 1 years when events or follow-up is based on downloaded data to address potential bias introduced by different follow-up times in the administrative and healthcare utilization databases
		3.6.1	Clarified that analysis of emergency room visits includes hospitalizations; additional statistical analysis of additional endpoints
		3.6.2	Corrected cut-and-paste error
		3.6.3	Clarified competing risk and other analyses for emergency room/hospital visits & chemotherapy duration; added statistical analysis of additional endpoints
		3.8	Added cancer diagnosis related covariates and clarified subgroup survival analysis
		4	Described source/timing of change to primary survival analysis
		5	Added reference of publication of secondary outcomes
3	September 30, 2022	Table of contents	Updated to reflect modified section name and page numbers
		2.1 (schema) 2.2.1 3.1.6 3.2.3 3.6.1 3.6.3	Further clarified the analysis of emergency room visits reverting to the originally intended analysis

		2.2.1 3.1.6 3.2.3 3.6.1 3.6.3	Removed download of claims (healthcare utilization) database and removed associated endpoints and statistical analysis
		3.5.3 3.6.3	Added analysis of time to death or deterioration at time of overall survival analysis in response to JAMA query

Table of contents

Glossary of abbreviations 5
1. Introduction 6
2. Study details 6
2.1. Study design 6
Study Calendar 7
2.2. Study objectives 10
2.2.1. Primary and secondary objectives 10
2.2.2. Qualitative and implementation objectives 10
2.3. Randomization and stratification criteria 10
2.4. Number of patients – initial sample size estimation 10
2.4.1. Accrual rate and accrual duration 10
2.4.2. Primary endpoint completion date for ClinicalTrials.gov reporting 10
2.5. Number of patients – sample size re-estimation 11
2.6. Power 11
2.7. Data safety and monitoring the study 11
3. Statistical methods for analysis 12
3.1. Data handling conventions 12
3.1.1. EORTC QLQ-C30 and CAHPS scoring algorithms 12
3.1.2. Data entry errors and potential outliers 12
3.1.3. Cleaning EORTC QLQ-C30, CAHPS, and other data 12
3.1.4. Missing data 13
3.1.5. Stratification errors 13
3.1.6. Integration of administrative (death) database 13
3.2. Types and time points of analyses 14
3.2.1. Interim analyses 14
3.2.2. Sample size re-estimation 14
3.2.3. Final analysis 14
3.3. Definition of populations for analyses 14
3.3.1. Modified intention-to-treat population for patient-level analysis 14
3.4. Study population description 15
3.4.1. Practice and patient disposition and exposure 15
3.4.2. Baseline demographic and other characteristics 15
3.5. Primary endpoint evaluation 15
3.5.1. Primary endpoint 15
3.5.2. Data set for primary endpoint 16
3.5.3. Analysis of primary endpoint 16
3.5.4. Sensitivity analyses for primary endpoint 16
3.6. Secondary and exploratory endpoint evaluation 17
3.6.1. Secondary and exploratory endpoints 17
3.6.2. Data sets for secondary endpoints 17
3.6.3. Analyses of secondary and exploratory endpoints 17
3.6.4. Sensitivity analyses for secondary endpoints 19
3.7. Safety endpoint evaluation 19
3.7.1. Safety endpoints 19
3.8. Subgroups 19
3.9. Software 20
4. Changes of analysis compared to study protocol 20
5. References 20

Glossary of abbreviations

AFT	Alliance Foundation Trials
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CONSORT	Consolidated Standards of Reporting Trials
CRA	Clinical research associate or assistant
DSMB	Data and Safety Monitoring Board
eCRF	Electronic case report form
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30
GED	General Educational Development test or Graduate Equivalency Degree or General Educational Diploma
GLM	General(ized) linear model
HRQL, HRQOL, HRQoL	Health-related QOL
IRB	Institutional Review Board
mITT	Modified intent-to-treat
PECD	Primary expected completion date
PRO	Patient-reported outcome
PRO-CORE	PRO-CORE is a consulting services and suite of tools for data collection including such as electronic patient surveys housed at the University of North Carolina
QALY	Quality-adjusted life years
QOL, QoL, QL	Quality of life
SAP	Statistical analysis plan
SD	Standard deviation
SDC	Statistics and Data Center
UNC	University of North Carolina
US	United States
USA	United States of America

1. Introduction

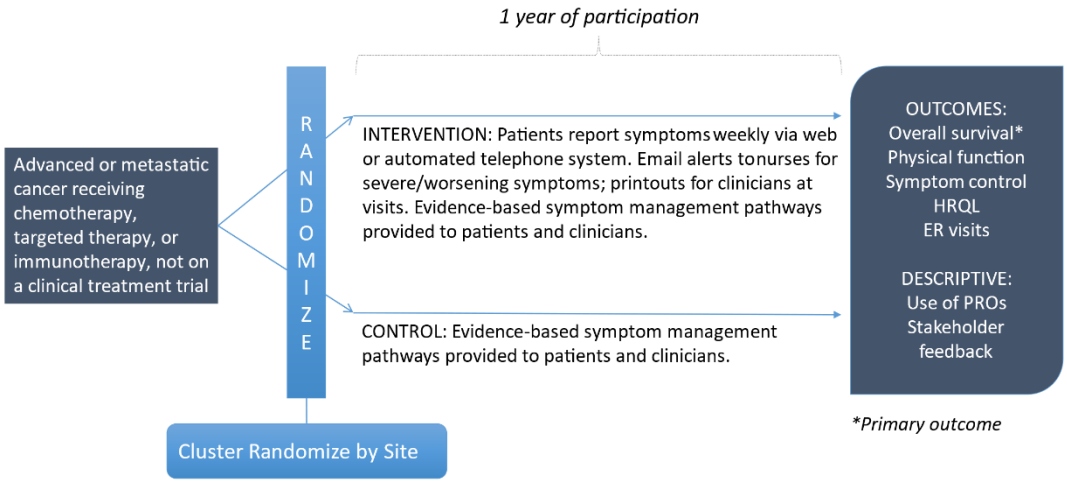
The SAP provides a detailed description of primary and secondary statistical analyses planned to be conducted within this trial at predefined time points. Subsequent and exploratory analyses are outside the scope of this document. Where possible, statistical analysis plans will be documented prior to initiation of subsequent statistical analyses.

2. Study details

2.1. Study design

This is a cluster randomized trial to evaluate the effects of systematic monitoring of symptoms via patient-reported outcome measures during routine cancer care delivery implemented at oncology practice sites in English, Spanish, or Mandarin Chinese-speaking adult cancer patients with advanced/metastatic cancer of any type (except leukemia or indolent lymphoma) receiving outpatient systemic cancer treatment.

PRO-TECT Schema:



The intervention is administered at the practice level. Data collection occurs at the patient, practice staff, and practice levels. Eligible patients will be approached at practices and asked to consent to participate. Patients who agree to consent will be registered through UNC's PRO-CORE system. See Study Calendar on next pages for planned assessments during this clinical trial.

Study Calendar

Control Sites Only

Source	Measure	Contents/Notes	Month of Patient Participation												Post	
			Base-line	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12 (or Off Study)*	*18*	24
Patient Reported (English, Spanish, Mandarin Chinese)	**P1. Patient Demographics**	Baseline characteristics	X														
	P2. Patient Quality of Life Questionnaire*	EORTC QLQ-C30 questions	X	X		X			X			X			X		
	P3. Patient Satisfaction Questionnaire*	CAHPS questions	X			X									X		
CRA Reported	**C1. Site Registration & Characteristics**	Site characteristics	Completed by CRA after a site has contracted to participate in the trial														
	C2. Patient Refusal to Participate/Ineligibility	Reason(s) and basic patient data	X														
	C3. Patient Registration	CRA must create/enter a unique patient ID; Some info requires abstracting medical record and input from patient or clinicians	X														
	C4. Patient Eligibility Checklist		X														
	C5. Additional Contact Information Form		X														
	C7. Patient Baseline Chart Abstraction Form	Info abstracted by CRA from participant's medical record	X														
	C9. Date of Death Form														X	X	X
	C12. Off Study Chart Abstraction Form**														X		
UNC	**UNC1. Site Training**	Details of startup meeting	X														

* ***The 3-month data collection is the key time point and is the most important date to have complete data collection***. *The patient questionnaires may be "bundled" together automatically by the PRO-Core software so it feels like a single longer questionnaire to participants. For Form P2, the timeframe is +/- 2 weeks for the month 1 form, and +/- 4 weeks for the months 3, 6, 9, and 12 forms. For Form P3 and Form P4, the timeframe for the month 3 and month 12 forms is +/- 4 weeks. If a participant does not complete a form within the specified time frame, the site CRA or UNC Coordinator should contact the patient to obtain this information. The site CRA and UNC Coordinator will work it out between them who will contact the patient.*

*** Window for completion is + 4 weeks.*

Intervention Sites Only

Source	Measure	Contents/Notes	Month of Patient Participation												Post	
			Base-line	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12 (or Off Study)*	*18*	*24*
Patient Reported (English, Spanish, Mandarin Chinese)	**Weekly PRO Survey – *Intervention Sites Only***	Symptom questions reported from home	X	X	X	X	X	X	X	X	X	X	X	X	X		
	P1. Patient Demographics	Baseline characteristics	X														
	P2. Patient Quality of Life Questionnaire*	EORTC QLQ-C30 questions	X	X		X			X			X			X		
	P3. Patient Satisfaction Questionnaire*	Questions about PRO system	X			X									X		
	P4. Patient PRO Feedback Booklet – *Intervention Sites Only**	CAHPS questions				X									X		
CRA Reported	**C1. Site Registration & Characteristics**	Site characteristics	Completed by UNC after a site has contracted to participate in the trial														
	C2. Patient Refusal to Participate/Ineligibility	Reason(s) and basic patient data	X														
	C3. Patient Registration	CRA must create/enter a unique patient ID; Some info requires abstracting medical record and input from patient or clinicians	X														
	C4. Patient Eligibility Checklist		X														
	C5. Additional Contact Information Form		X														
	C6. Missed Weekly Patient PRO Survey – *Intervention Sites Only*§	Info collected from patients by site CRA (or assisted by UNC)	Collected if participant misses a scheduled *Weekly PRO Survey*. Reason for missed survey should be selected.														
	C7. Patient Baseline Chart Abstraction Form	Info abstracted by CRA from medical record	X														
	C8. Patient Contact Log for Missed PRO Survey – *Intervention Sites Only*§	Info collected from patients by site CRA (or assisted by UNC)	Completed after successful or unsuccessful attempts to contact participants to collect information for *Form C6*.														
	C9. Date of Death Form	Info abstracted by CRA from medical record													X	X	X
	C10. CRA Perspectives– *Intervention Sites Only*§	Questions for CRAs about PRO system	To be completed after study has been open at site for at least 6 months.														
	C11. Nursing Alert Response Form– *Intervention Sites Only*	CRA obtains responses from clinical nurse who got the alert	Collected within 72 hours of each nursing alert notification, to elicit actions taken by clinical nurse in response to the alert														

	C12. Off Study Chart Abstraction Form**	Info abstracted by CRA from medical record												X		
	Printed PRO Report	Patients' symptoms	Printed for oncologist and nurse at clinic visits.													
Nurse Reported	**N1. Nurse Perspectives– *Intervention Sites Only*§**	Questions about PRO system	To be completed after study has been open a site for at least 6 months.													
Oncologist Reported	**Onc1. Physician Response Form**	Questions about PRO Report Usage	To be completed after study has been open a site for at least 6 months.													
UNC	**UNC1. Site Training**	Details of startup meeting	X													

* ***The 3-month data collection is the key time point and is the most important date to have complete data collection****. The patient questionnaires may be "bundled" together automatically by the PRO-Core software so it feels like a single longer questionnaire to participants. For Form P2, the timeframe is +/- 2 weeks for the month 1 form, and +/- 4 weeks for the months 3, 6, 9, and 12 forms. For Form P3 and Form P4, the timeframe for the month 3 and month 12 forms is +/- 4 weeks. If a participant does not complete a form within the specified time frame, the site CRA or UNC Coordinator should contact the patient to obtain this information. The site CRA and UNC Coordinator will work it out between them who will contact the patient.*

*** Window for completion is + 4 weeks.*

§ *To be completed after the study has been open at a site for at least 6 months. The form should be collected within a week of this time point, but there is no expiration on the timeframe for collecting these up through study closure.*

‡ *The site CRA and UNC Coordinator will work it out between them who should be contacting their site's participants who do not complete the Weekly PRO Survey on time (within 24 hours) for backup/reminder/questions. This information should be collected as soon as possible but can be collected up until the day*

2.2. Study objectives

2.2.1. Primary and secondary objectives

- Determine whether systematic monitoring of symptoms via patient-reported outcome measures during routine cancer care delivery improves meaningful clinical outcomes including survival, health utilities/quality-adjusted life years (QALYs), quality of life, symptom control, emergency room visits, and patient satisfaction with care. Feasibility, individual symptoms, other functioning scales, alert notifications, and financial toxicity as reported by the patient will also be explored.

2.2.2. Qualitative and implementation objectives

- Elicit perspectives from patients, CRAs, and clinicians about effort, benefits, and burden of patient self-reporting of symptoms with alerts and reports to clinicians.
- Identify barriers, facilitators, and strategies used by practices to integrate PROs into clinical workflow through interviews, questionnaires, and selected site visits, including impact of patient characteristics such as race, ethnicity, computer experience, or educational background.
- Obtain perspectives of stakeholders about PROs through debriefings at study completion.
- Evaluate financial impact of patient self-reporting.

The analysis of these objectives will be contained in separate SAPs.

2.3. Randomization and stratification criteria

Practices will be randomly assigned to each arm in a 1:1 ratio by the AFT Statistics and Data Center based at the Mayo Clinic, using permuted block randomization with random block size of 2 or 4 stratified by rural vs. urban location. The randomization sequences (one for each stratum) will remain concealed and arm assignments will only be generated and revealed one at a time as practices are registered by the UNC Coordinator.

2.4. Number of patients – initial sample size estimation

Initial target sample size was 1,000 patients from up to 50 U.S. practices. This was later amended to 1,200 patients from up to 50 (+/-5) U.S. practices.

2.4.1. Accrual rate and accrual duration

Accrual period is expected to be approximately 3 years. The intervention period is 12 months. The practices will follow patients for 24 months for overall survival. Annual national administrative database downloads will be used to capture overall survival status of patients to capture patient death status after this period.

2.4.2. Primary endpoint completion date for ClinicalTrials.gov reporting

At study activation, this study will have been registered within the ClincialTrials.gov website. The Primary and Secondary Endpoints (i.e., “Outcome Measures”) along with other required information for this study will be reported on ClinicalTrials.gov.

- For purposes of timing of the Results Reporting, the initial estimated completion date for the Primary Endpoint of this study is 5 years after the study opens to accrual (3 years for accrual and 2 years of practice follow-up).

- The definition of "Primary Endpoint Completion Date" (PECD) for this study is the date of national administrative database download to support the primary overall survival analysis.

2.5. Number of patients – sample size re-estimation

There is no additional planned sample size re-estimation in this study.

2.6. Power

For overall survival, the trial's initial sample size of 1,000 patients from 50 U.S. practices provided 80% power for a hazard ratio of 0.76 (based on the prior single-center randomized controlled trial) which is considered clinically meaningful (Sobrero AF, et al. 2015; Ellis LM, et al. 2014) using a two-sided alpha=0.05/2 log-rank test with 521 events observed during the observation period, computed using the formula by Xie and Waksman, 2003, with an intracluster correlation coefficient of 0.001 (estimated from the 10 largest legacy Alliance trials involving 12,717 total patients). This sample size also provided 95% power to measure the established clinically meaningful difference of 0.37 standard deviations on the QLQ-C30 Physical Function Scale based on a prior single-center randomized controlled trial (~9 points on the 100-point QLQ-C30 scale) between randomization groups using a two-sided alpha=0.05/2 t-test assuming an intracluster correlation coefficient of 0.055 (Adams G, et al. 2004), and assuming that 85% of patients are evaluable for the primary analysis at the 3-month time point.

In order to optimize power, the sample size was subsequently amended to 1,200 patients from 50 (+/-5) U.S. practices. Power for comparing physical functioning increased to 96-97% and power remained 80% for comparing overall survival under the assumption that statistical analysis would be conducted after 522 events. Each analysis would be undertaken with a two-sided alpha=0.05/2.

Finally, after an additional amendment to increase type I error for the overall survival analysis, with a total of 1,200 patients at 52 practices nationally, there was at least 90% power for a hazard ratio of 0.76 (based on the prior single-center randomized controlled trial) which is considered clinically meaningful using a two-sided alpha=0.05 log-rank test with 576 observed events, computed using the formula by Xie and Waksman, 2003, with an intracluster correlation coefficient of 0.001 (estimated from the 10 largest legacy Alliance trials involving 12,717 total patients). This power calculation further assumes drop-out of 150 patients in the first 2.5 years.

2.7. Data safety and monitoring the study

This study is monitored by the IRB and study team. No formal DSMB will be employed because no safety concerns are associated with administering questionnaires (i.e., the intervention) to patients.

3. Statistical methods for analysis

3.1. Data handling conventions

3.1.1. EORTC QLQ-C30 and CAHPS scoring algorithms

The EORTC QLQ-C30 contains 30 questions that assess various domains of HRQL. Scoring for the EORTC QLQ-C30 can be found in the Scoring Manual available from the EORTC. The standard scoring of the EORTC QLQ-C30 generates scales or item scores. Gundy CM, et al. (2012) described additional composite scales. In planned analyses, Physical Functioning is based on Q1-5; Symptom Burden ("Control") is based on 8 symptom scales/items (Q8-19); Health-Related Quality of Life is based on 5 functional scales and 8 symptom scales/items (Q1-27); Fatigue is based on Q10, Q12, & Q18; Nausea and Vomiting on Q14 & Q15; Pain on Q9 & Q19; Dyspnea on Q8; Insomnia on Q11; Appetite Loss on Q13; Constipation on Q16; and Diarrhea on Q17. Analysis of QALYs will additionally compute health utilities from EORTC QLQ-C30 data using the algorithm for U.S. patients published by Revicki DA, et al. (2021).

Items from the CAHPS survey were additionally administered. These items will be analyzed as individual items and no scoring algorithm will be applied to combine items into scale scores.

3.1.2. Data entry errors and potential outliers

Electronic CRF (eCRF) fields may contain potential outliers or suspected data entry errors due to inconsistencies with other entered data values. If such values are identified, the data management team will evaluate the data and communicate with the practice for potential resolution. Sensitivity analyses excluding these measurements may be carried out to evaluate the influence of such values. Values found to be incorrect due to data entry error after data freeze may be hard-coded in the data preparation program.

3.1.3. Cleaning EORTC QLQ-C30, CAHPS, and other data

Attempts will be made to obtain missing date information for primary and secondary endpoints including for evaluating whether questionnaires fall within acceptable time frames for inclusion in statistical analysis.

Dates of patient questionnaire completion will be reviewed to assess whether questionnaires were entered at the correct time points in the database. Note that UNC's PRO-CORE system will automatically date and time stamp electronically administered questionnaires. If duplicate questionnaires are suspected to be entered/completed, questionnaires are suspected to be entered at erroneous time points, or other data entry or completion errors pertaining to the questionnaires are suspected, the data management team will communicate with the practice for potential resolution. Data entry errors pertaining to the questionnaires found after data freeze may be hard-coded in the data preparation program.

After data cleaning, the following rules will be applied to select patient questionnaire data for inclusion in analysis of the EORTC QLQ-C30 and CAHPS data:

- Questionnaires will be included in the time point that was intended for administration
- If a patient completes the same questionnaire more than once at a given time point, the first questionnaire that was completed within the given time point will be retained

- If the patient ends the study early and completes an "end of study" questionnaire prior to 12 months, that questionnaire will be included in the nearest corresponding time point

In eCRF items with associated "other" free text fields, the free text values will be reviewed. If free text values are identified which match an available and specific response option, the data management team will communicate with the practice for potential updating of the eCRF field. Such values that are identified after data freeze may be hard-coded in the data preparation program.

3.1.4. Missing data

We first plan to minimize missing data prospectively through the use of central data collection in PRO-CORE at UNC with prospective monitoring of data consistency and completeness. Prospectively, we also plan regular contact with practices to identify missing data problems early. Analytically, for the primary overall survival analysis we are employing a large national database to augment practice-reported death data to ensure that reported deaths are as complete as possible. If any questions arise regarding status of a patient, the last known date alive will be used for censoring. For patient questionnaire-based analyses, we will use model-based approaches to incorporate all available data and to minimize the impact of missing data. In the event of a high number of missing questionnaires, we will compare baseline patient characteristics between patients who do and don't complete questionnaires for a given patient questionnaire-based analysis. If selection bias is a concern, we will employ multiple imputation in sensitivity analyses.

3.1.5. Stratification errors

If a stratification error is noted after a practice is randomized, there will be no change in the randomization system. The stratification factor on the eCRF will collect the correct stratification factor values. The stratification factor value as recorded in the eCRF will be used for the analysis.

3.1.6. Integration of administrative (death) database

Incorporation of deaths from administrative source downloads for the overall survival analysis will be undertaken in a blinded fashion such that randomized arm and eCRF fields unique to a single arm will not be considered. Note that a decision was made in 2022 to not pursue purchase of claims (healthcare utilization) data so analyses related to such data were removed from the SAP. For the overall survival analysis, the general procedure will include comparison of a death in the administrative database to a patient's eCRF data to assess whether the death should be included in the analysis data set. Timing of the death relative to other eCRF variables including site-reported death date and strength of match will be used to assess whether the death should be included. Sites will be queried for discrepancies greater than 7 days between eCRF-reported death date and administrative database death date and for patients in which a death appears in only one data source. The eCRF data will be prioritized over administrative data in the event of uncertainty. Deaths and emergency room visits after study enrollment will be integrated regardless of time since study enrollment. Censoring to account for potential imbalance of follow-up will be handled through sensitivity analyses.

3.2. Types and time points of analyses

3.2.1. Interim analyses

There is no planned interim analysis in this study.

3.2.2. Sample size re-estimation

There is no additional planned sample size re-estimation in this study.

3.2.3. Final analysis

Analysis of intervention adherence rates, EORTC QLQ-C30 outcomes, and other data during the intervention phase will occur when all patients have been followed for 12 months (or ended participation prior to 12 months).

Overall survival analysis will be undertaken when at least the required number of survival events have been observed. Survival data will be derived from both eCRFs (practice reported) and through linkage to an administrative national database, therefore statistical analysis may not be able to be undertaken at the exact specified number of events and instead would be timed according to data download (i.e., the number of deaths may be contingent on findings from downloaded data). Note that national death databases are delayed in releasing death information, and will require collection of identifying information from sites for data requests, so we anticipate the first analytical download of data to occur in 2022. Annual downloads will occur after that until the required number of deaths have been observed. Annual downloads may continue depending on funding availability after primary analysis to acquire longer term follow-up. Statistical analysis will use all deaths recorded in the study database as well as downloaded data at time of analysis including deaths in excess of the number needed to trigger the planned analysis. The overall survival analysis will censor events and/or last follow-up date at 2 years (i.e., 730 days plus 6 weeks) to standardize the follow-up time across patients in both arms.

The analysis of emergency room visits may employ data derived from eCRFs (practice reported) and/or through linkage to an administrative national database (database employed for survival data to assist in censoring). Like the overall survival analysis, analysis of these data will occur when data become available. The first analysis will occur when administrative data from 2020 becomes available (or at time of overall analysis if administrative data download is not pursued) and additional downloads will be undertaken as necessary for data completeness and depending on funding availability. Like the overall survival analysis, the analysis will censor events and/or last follow-up date at 1 year (i.e., 365 days plus 6 weeks) to standardize the follow-up time across patients in both arms.

3.3. Definition of populations for analyses

3.3.1. Modified intention-to-treat population for patient-level analysis

We will use a modified intention-to-treat (mITT) principle to define the population used for the analysis of the primary endpoint, which is based on data collected at the patient level. The mITT population will be comprised of all patients who consent to participate and are registered through UNC by any randomized practice. We will exclude any patient who discontinues their cancer treatment on or prior to their date of registration. Patients will be analyzed according to the practice in which they are registered, and practices will be analyzed according to the intervention group to which they were randomized. All available data will be

included in each analysis. The number of patients included in a given analysis will depend on data availability (e.g., analysis of EORTC QLQ-C30 at 12 months will include all patients who completed the necessary EORTC QLQ-C30 items to produce each scale score at 12 months).

3.4. Study population description

3.4.1. Practice and patient disposition and exposure

Practice randomization and practice status (eligibility, randomization, and discontinuation from study) as well as patient enrollment and patient status (eligibility and discontinuation from study) will be listed and summarized by randomized arm. A CONSORT diagram will be generated following conventions recommended for cluster randomized trials (Campbell MK, et al. 2012).

3.4.2. Baseline demographic and other characteristics

At the practice level, the stratification factor (practice location [rural vs urban]) will be summarized by randomized arm. At the patient level, the following will be summarized by randomized arm: demographic variables (age [median, minimum, maximum], sex, race, and ethnicity), weekly PRO survey mode (for intervention arm only), education level, employment status, practice location (stratification factor described at the patient level instead of the practice level), prior cell phone use, prior internet use, prior email use, difficulty paying monthly bills, cancer type, line of systemic cancer treatment, days since first diagnosis to metastatic cancer, and days since metastatic cancer to study enrollment.

We do not anticipate formally comparing patient-level baseline characteristics between randomized arms to assess for balance. If comparison becomes necessary during analysis, we will use a generalized linear mixed model using the appropriate link function based on the distribution of the variable with a random practice intercept term to account for clustering within practice.

3.5. Primary endpoint evaluation

3.5.1. Primary endpoint

At time of study activation, the protocol defined two primary endpoints: Physical function at 3 months as measured by the EORTC QLQ-C30 Physical Function Scale; and overall survival. The two associated hypotheses tested for superiority of the intervention arm:

Hypothesis (physical functioning): Within oncology practices that are randomized to systematic monitoring of symptoms via patient-reported outcome measures, physical functioning 3 months after registration of patients will be higher as compared to patients at oncology practices randomized to usual care.

Hypothesis (overall survival): Within oncology practices that are randomized to systematic monitoring of symptoms via patient-reported outcome measures, overall survival of patients will be higher as compared to patients at oncology practices randomized to usual care.

After amending the protocol, overall survival was identified as the primary endpoint. Physical functioning as previously defined will be considered as a key secondary endpoint. Hypotheses remain unchanged.

3.5.2. Data set for primary endpoint

For the main analysis of the primary endpoint, the mITT population defined in SAP Section 3.3.1 will be used.

3.5.3. Analysis of primary endpoint

The primary analysis of overall survival will employ a Cox regression with cancer diagnosis related covariates (e.g., line of systemic cancer treatment, days since first diagnosis to metastatic cancer, days since metastatic cancer to study enrollment) and a random effect to account for site clustering. All deaths will be included in the analysis and patients without observed deaths will be censored on the last date known alive using all available eCRF and administrative database data. Final presented model may omit nonsignificant covariates. Additional details of this analysis were documented in an updated version of this SAP between analysis of secondary outcomes and initiation of primary overall survival analysis to account for necessary modifications for carrying out quality control of administrative data, defining events, and implementing censoring rules. The hazard ratio over time will be plotted.

At the time of the overall survival analysis, we will conduct an updated analysis of key patient-reported outcomes (EORTC QLQ-C30 physical function, symptom burden, and HRQL) combined with overall survival data. Time to death or deterioration in each domain will be defined as time to first 10-point decline from baseline or death due to any cause within 1 year (i.e., 365 days plus 6 weeks). Cox regression will include the same covariates and random effect as the overall survival model previously described. Patients without observed 10-point declines or deaths will be censored at date of last questionnaire completed. A frequency table will be used to explore types of events (death or deterioration) contributing to each analysis. A sensitivity analysis will consider time to first 10-point decline in each domain without considering death as an event.

Additional analysis will also include comparison of health utilities and QALYs between arms using health utilities computed based on the EORTC QLQ-C30. Patient QALYs will be computed using the area-under-the-curve approach (without discounting) and will include all data through a consistent time point to avoid bias related to censoring, such as the follow-up of the last consented patient (i.e., the earliest censored patient). A population-based approach will be used such that the area-under-the-curve of a quality-adjusted survival curve (group or arm mean health utility based on a mixed model multiplied by each patient's survival) is the mean quality-adjusted survival for the population. Mean quality-adjusted survival will be compared between arms using a bootstrap approach.

3.5.4. Sensitivity analyses for primary endpoint

Supplemental analyses will include model-based analysis with subgroupings based on patient and practice characteristics (see SAP Section 3.8). Also see SAP Section 3.1.4 for sensitivity analyses to assess the impact of missing data. Lastly, the overall survival analysis will be repeated with censoring of events and/or last follow-up date to 1 year (i.e., 365 days plus 6 weeks) to standardize the follow-up time across patients in both arms.

3.6. Secondary and exploratory endpoint evaluation

3.6.1. Secondary and exploratory endpoints

As described above, the key secondary endpoint is physical function at 3 months as measured by the EORTC QLQ-C30 Physical Function Scale. Additional secondary endpoints include: EORTC QLQ-C30 symptom burden (control) score at 3 months, and EORTC QLQ-C30 HRQL score at 3 months.

Additional endpoints include: EORTC QLQ-C30 appetite loss, constipation, diarrhea, dyspnea, fatigue, insomnia, nausea/vomiting, and pain score at 3 months; EORTC QLQ-C30 function scales at 3 months, patient adherence to the intervention (i.e., completion of the weekly symptom surveys); CAHPS item scores at 3 months; emergency room visits; alert notification; and financial toxicity.

3.6.2. Data sets for secondary endpoints

For analysis of secondary endpoints, the mITT population defined in SAP Section 3.3.1 will be used.

3.6.3. Analyses of secondary and exploratory endpoints

Mean change from baseline in physical function, symptom control, and HRQL at 3 months will be compared between arms using a linear combination of parameters from a general linear mixed model. Each model will include all available data from all time points (months 0, 1, 3, 6, 9, and 12). Fixed effects will include arm, time point, cancer type, and arm-by-time point interaction. A random practice intercept term will be included to account for clustering by practice. Repeated observations by patient will be modeled using compound symmetric correlation structure over time. Such values as the mean change from baseline at 3 months by arm, and difference in mean change from baseline at 3 months between arms will be estimated with confidence intervals based on the mixed model. Comparisons at other time points will also be carried out and graphically displayed using mean plots.

To supplement comparison of means, a responder analysis will also be employed. Patients who complete the QLQ-C30 at baseline and at 3 months will be categorized as improving on each outcome (physical function, symptom control, HRQL) if their score increased by 5 or more points from baseline; worsening if their score decrease by 5 or more points from baseline; and otherwise as stable. The selection of a 5-point change on the 100-point QLQ-C30 scale was selected as clinically meaningful based on work by Cocks K, et al. 2012. The proportion of patients with improvement, stability, or worsening will be compared between arms at 3 months using a cumulative logistic regression model with fixed effects for arm and cancer type and a random practice intercept term to account for clustering by practice. Similar to the mean comparisons, the responder analysis will be carried out at other time points as needed to supplement that primary analysis at 3 months.

See Section 3.5.3 for statistical analysis combining physical function, symptom control, and HRQL with overall survival data to be conducted at time of overall survival analysis.

Patient adherence (feasibility) to the intervention (i.e., completion of the weekly symptom surveys) is defined at each week as the proportion of participating patients completing a survey divided by the number of participating patients who are expected to complete a survey. Completion will be computed at each week and overall.

CAHPS items will be analyzed descriptively using frequency and relative frequency of each response option. Items will be described at each time point.

Analysis of emergency room visits will employ similar approaches as the overall survival analysis, however, using Fine-Gray competing risk regression with death as a competing event instead of Cox regression. The same covariates as the overall survival analysis will be incorporated into the competing risk analysis. Number of emergency room visits will be explored using a marginal means/rates model. Alternative modeling approach may be used depending on the observed count/pattern of events. The analysis of emergency room visits will censor events and/or last follow-up date at 1 year (i.e., 365 days plus 6 weeks) to standardize the follow-up time across patients in both arms. Additional details of these analyses were documented in an updated version of this SAP between date of first administrative database download and initiation of analysis of these outcomes to account for necessary modifications for carrying out quality control of administrative data, defining events and competing risks, and implementing censoring rules. SAP was finalized prior to initiation of statistical analysis.

Additional analyses include statistical analysis of other QLQ-C30 scales (individual symptoms and function scales), alert notifications, and financial toxicity as reported by the patient. The analysis of other QLQ-C30 scales will be similar to the mixed models, graphics, and responder analyses carried out for QLQ-C30 physical function, symptom burden, and HRQL. Covariates may be modified to include cancer diagnosis related variables (e.g., line of systemic cancer treatment, days since first diagnosis to metastatic cancer, days since metastatic cancer to study enrollment) like the overall survival analysis.

The number of alerts, reasons for the alerts, patient and nurse opinions regarding the urgency of an alert, and immediate nurse actions in response to an alert will be tabulated descriptively. To predict the nurse's opinion regarding the urgency of an alert, a generalized linear mixed model with a logit link will be employed with fixed effects for patient characteristics at baseline and information from the weekly PRO surveys (e.g., patient-reported symptoms, worsening of each symptom from the prior week, and prior receipt of an alert for each symptom) and a random patient intercept. To assess the added benefit of the weekly PRO surveys, results will be compared to those from a generalized linear mixed model with baseline patient characteristics (only) as fixed effects. Receiver operating characteristic (ROC) curves will be plotted along with computed areas-under-the-curve. Sensitivity, specificity, positive predictive value, and negative predictive value will be estimated with 95% Wald confidence intervals for various cutpoints (e.g., cutpoint for likelihood of an alert being urgent that achieves 80% sensitivity).

In March 2019, a monthly financial toxicity question was added to the weekly PRO surveys completed by patients at PRO practices. Financial difficulties were also measured for all patients (both arms, entire study) using a single item from the EORTC QLQ-C30. Baseline financial difficulties (as measured by the EORTC QLQ-C30) will be tabulated overall and by arm, and the relationship between financial difficulties and patient characteristics will be explored using chi-square tests for categorical variables and Wilcoxon Rank Sum tests for continuous variables. The proportion of patients with any new/worsening financial difficulties (as measured by the EORTC QLQ-C30) will be compared between arms using a generalized linear mixed model with fixed effects for arm and a random practice intercept term to account for clustering by practice. This analysis will be completed using all patients with data, as well as in the subset of patients who participated during the time in which a financial toxicity question was administered in the PRO arm and the subset of patients who participated entirely before the time in which a financial toxicity question was administered in the PRO arm.

3.6.4. Sensitivity analyses for secondary endpoints

As described above, supplemental analyses will include mean and responder comparisons at time points other than 3 months. For the responder analysis, a 10-point change will also be applied to ensure results remain consistent with a higher threshold. Finally, impact of the COVID pandemic will be explored by repeating key between-arm analyses within the subset of patients enrolled prior to December 1, 2019 (date selected to ensure that the 3-month time point occurred prior to widespread COVID pandemic impacts on clinical practices in the US). Outcomes of patients enrolled after December 1, 2019, may also be tabulated and compared to outcomes of patients enrolled prior to December 1, 2019.

3.7. Safety endpoint evaluation

3.7.1. Safety endpoints

As the intervention is the administration of questionnaires, no safety data will be collected as part of this study. See Section 3.6.3 for analysis describing patient adherence to the intervention.

3.8. Subgroups

The following baseline patient characteristics will be used to perform subgroup analyses:

Age (<60 versus ≥60)
Gender (female versus male)
Race (white versus non-white)
Ethnicity (Hispanic versus non-Hispanic)
Education status (high school graduate/GED or less versus some college or more)
Employment status (working versus not currently working)
Marital status (married/partnered versus other)
Prior computer use (rarely or less versus sometimes or more)
Prior email use (rarely or less versus sometimes or more)
Prior internet use (rarely or less versus sometimes or more)
Practice location (rural versus urban)
Cancer type (thoracic, breast, colorectal, genitourinary, gynecologic versus other)
Line of systemic cancer treatment (1^{st} versus 2^{nd} versus $\geq 3^{rd}$)
Days since first diagnosis to metastatic cancer (continuous variable)
Days since metastatic cancer to study enrollment (continuous variable)

Covariates may not be omitted from analyses or groupings (combinations of levels) may be adapted based on observed sample size. Additional covariates may be added but will be considered as post-hoc. Continuous covariates may be split by the observed median or tertiles.

Subgroup analysis will be carried out for the primary and key secondary endpoints (including but not limited to overall survival and EORTC QLQ-C30 Physical Function Scale at 3

months). Additional subgroup analyses may be carried out for other time points and/or other EORTC QLQ-C30 scales as indicated though care will be taken to note inflation of type I error when incorporating additional analyses. To carry out subgroup analyses, a general linear mixed model within each level of the subgroup variable will be fit using data from all time points (months 0, 1, 3, 6, 9, and 12), and effect of the intervention at month 3 will be tested similar to the primary analysis using a linear combination of model parameters. Such values as the mean change from baseline at 3 months by arm, and difference in mean change from baseline at 3 months between arms will be estimated with confidence intervals based on these mixed models. Next, a general linear mixed model will be fit including all patients and including fixed effects for arm, time point, and the given subgrouping variable, as well as pairwise interactions between arm and time point, arm and subgroup variable, and time point and subgroup variable. Higher order interactions will initially be included in models but may be omitted if lacking statistical significance. A random practice intercept term will account for clustering by practice. Repeated observations by patient will be modeled using compound symmetric correlation structure over time. The Type 3 test of the interaction effect between arm and the given subgroup variable will be used to formally test for statistically significant subgroup effect. The value(s) of the coefficient(s) on the interaction effect between arm and subgroup will be used to assess the magnitude of the interaction. A similar approach as described above but using Cox regression will be used for subgroup analysis of overall survival. Regression will be carried out within each level of a subgroup variable and carried out across levels with a subgroup-by-arm interaction term to formally test for statistically significant subgroup effect.

3.9. Software

All analyses will be done using SAS version 9.4 or higher, or R version 3.5.1 or higher by members of the Alliance Statistics and Data Center. If necessary (e.g., if statistical tests are not offered in the software used) analyses will be carried out using other software.

4. Changes of analysis compared to study protocol

Existing subgroups were modified and additional subgroups were added to this SAP relative to the subgroupings listed in the protocol. These changes were made based on early descriptive analysis reported by Basch E, et al. 2020, and implemented prior to initiation of the primary analysis. The approach to the primary survival analysis was modified based on observed descriptive differences in cancer diagnosis related covariates between arms in Basch E, et al. 2022, and implemented prior to initiation of the primary survival analysis.

5. References

Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. J Clin Epidemiol. 2004 Aug;57(8):785-94.

Basch E, Schrag D, Henson S, Jansen J, Ginos B, Stover AM, Carr P, Spears PA, Jonsson M, Deal AM, Bennett AV, Thanarajasingam G, Rogak LJ, Reeve BB, Snyder C, Bruner D, Cella D, Kottschade LA, Perlmutter J, Geoghegan C, Samuel-Ryals CA, Given B, Mazza GL, Miller R, Strasser JF, Zylla DM, Weiss A, Blinder VS, Dueck AC. Effect of Electronic Symptom Monitoring on Patient-Reported Outcomes Among Patients With

Metastatic Cancer: A Randomized Clinical Trial. JAMA. 2022 Jun 28;327(24):2413-2422. PMID: 35661856; PMCID: PMC9168923.

Basch E, Stover AM, Schrag D, Chung A, Jansen J, Henson S, Carr P, Ginos B, Deal A, Spears PA, Jonsson M, Bennett AV, Mody G, Thanarajasingam G, Rogak LJ, Reeve BB, Snyder C, Kottschade LA, Charlot M, Weiss A, Bruner D, Dueck AC. Clinical utility and user perceptions of a digital system for electronic patient-reported symptom monitoring during routine cancer care: findings from the PRO-TECT trial. JCO Clin Cancer Inform. 2020 Oct;4:947-957. PMID: 33112661; PMCID: PMC7768331.

Campbell MK, Piaggio G, Elbourne DR, Altman DG; CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012; 345:e5661. PubMed PMID: 22951546.

Cocks K, King MT, Velikova G, de Castro G Jr, Martyn St-James M, Fayers PM, Brown JM. Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. Eur J Cancer. 2012 Jul;48(11):1713-21. PMID: 22418017.

Ellis LM, Bernstein DS, Voest EE, Berlin JD, Sargent D, Cortazar P, Garrett-Mayer E, Herbst RS, Lilenbaum RC, Sima C, Venook AP, Gonen M, Schilsky RL, Meropol NJ, Schnipper LE. American Society of Clinical Oncology perspective: Raising the bar for clinical trials by defining clinically meaningful outcomes. J Clin Oncol. 2014 Apr 20;32(12):1277-80.

Gundy CM, Fayers PM, Groenvold M, Petersen MA, Scott NW, Sprangers MA, Velikova G, Aaronson NK. Comparing higher order models for the EORTC QLQ-C30. Qual Life Res. 2012 Nov;21(9):1607-17. PMID: 22187352; PMCID: PMC3472059.

Sobrero AF, Pastorino A, Sargent DJ, Bruzzi P. Raising the bar for antineoplastic agents: how to choose threshold values for superiority trials in advanced solid tumors. Clin Cancer Res. 2015 Mar 1;21(5):1036-43.

Revicki DA, King MT, Viney R, Pickard AS, Mercieca-Bebber R, Shaw JW, Müller F, Norman R. United States Utility Algorithm for the EORTC QLU-C10D, a Multiattribute Utility Instrument Based on a Cancer-Specific Quality-of-Life Instrument. Med Decis Making. 2021 May;41(4):485-501. PMID: 33813946.

Xie T, Waksman J. Design and sample size estimation in clinical trials with clustered survival times as the primary endpoint. Stat Med. 2003; 22(18):2835-46.