STATISTICAL ANALYSIS PLAN
for
PROTOCOL ACTIVITY 4
Pragmatic stepped-wedge cluster randomized trial

SIMPRO Research Center: Integration and Implementation of PROs

for Symptom Management in Oncology Practice.

VERSION NUMBER: 2.0

DATE: *January 9, 2023*

Sponsor and Coordinating Center:

Dana-Farber/Harvard Cancer Center (DF/HCC)
Department of Medical Oncology
Coordinating Center PI and Technology PI: Michael Hassett
Dana-Farber Cancer Institute
Boston, MA 02215
Email Address: Michael_Hassett@dfci.harvard.edu
Telephone Number: 617-632-4587

CO-STUDY CHAIRS:

Deborah Schrag MD, MPH
Chair, Department of Medicine
Memorial Sloan Kettering Cancer Center
New York City, NY 10065

Raymond Osarogiagbon MD
Director, Thoracic Oncology Research Group
Baptist Memorial Hospital
Covington, Tennessee 38019

Sandra Wong MD
Chair, Department of Surgery
Dartmouth College
Lebanon, NH 03756

STUDY STATISTICIAN:

Hajime Uno, PhD
Dana-Farber Cancer Institute
Boston, MA 02215

FUNDING:

National Cancer Institute, UM1CA233080.

SIGNATURE PAGE

Protocol Title:	SIMPRO Research Center: Integration and Implementation of PROs for Symptom Management in Oncology Practice.
Sponsor Protocol Number:	DF/HCC IRB #18-986R/18-734
sIRB Protocol Number:	WIRB Tracking #20182593; Study #1248093
SAP Version:	Version 2.0
Date:	January 9, 2023

Michael Hassett, MD, MPH Study Chair, Coordinating Center PI	Date
Deborah Schrag, MD, MPH Co-Chair, Co-PI	Date
Raymond Osarogiagbon, MD Co-Chair, Co-PI	Date
Sandra Wong, MD Co-Chair, Co-PI	Date
Hajime Uno, PhD Study Statistician	Date

RECORDS ON REVISIONS

Version	Date for creation or modification	Person in charge	Remarks
Ver. 1.0	July 20, 2023	Hajime Uno	
Ver. 2.0	January 9, 2024	Hajime Uno	

ACRONYMS USED THROUGHOUT SAP:

AD	Absolute Difference
ED	Emergency Department
EDTR	Emergency Department Treat/Release
EHR	Electronic health record
ePRO	Electronic patient-reported outcomes
eSyM	Electronic symptom management system
eSyM+	eSyM intervention condition
eSyM-	eSyM control condition
GI	Gastrointestinal
GLMM	Generalized Linear Mixed-effects Model
Gyn	Gynecologic
ICC	Intra Class Correlation
IRD	Incidence Rate Difference
IRR	Incidence Rate Ratio
MO	Medical Oncology
OR	Odds Ratio
SAP	Statistical Analysis Plan
SASS	Self-efficacy, Attainment of information needs, Symptom burden, and Satisfaction (Research questionnaire)
Surg	Surgery
SW-CRT	Stepped Wedge Cluster Randomized Trial
THOR	Thoracic

Table of Contents

Table of Contents

1. Introduction

This document describes the statistical analysis plan for the multisite pragmatic stepped-wedge cluster randomized trial (SW-CRT) specified as "Activity 4" in the study protocol *"SIMPRO Research Center: Integration and Implementation of PROs for Symptom Management in Oncology Practice."*

The study protocol consists of four activities. The overall research goals are: (1) to create and refine eSyM, a reporting and management system that integrates ePROs with the EHR; (2) to evaluate the impact of eSyM on patient outcomes, treatment delivery, and healthcare system utilization using a pragmatic cluster randomized study design; and (3) to undertake a systematic, deliberative approach to implementation to allow for the identification of barriers and facilitators that contribute to the adoption and sustainability of eSyM in routine oncology care.

Table 1 shows the Aims for the entire study. The SW-CRT conducted in Activity 4 addresses Aims 2 and 3. *This document covers the statistical analysis plan for Aim 2 and Aim 3 only.*

Table 1: Aims of the project: "SIMPRO Research Center: Integration and Implementation of PROs for Symptom Management in Oncology Practice."
Aim 1: Adapt existing ePRO symptom management systems and integrate them into the EHR and routine clinical workflow at six health systems. Specifically: Aim 1a. Obtain patient, clinician, staff and leadership input on ePRO form and function Aim 1b. Refine the content and algorithms for self-management, alerts, and feedback Aim 1c. Develop ePRO training materials for patients, clinicians, and staff Aim 1d. Pilot an ePRO symptom manager at test and study sites and prepare an implementation strategy
Aim 2: Determine the effectiveness of eSyM (an EHR-integrated ePRO symptom management system) on health outcomes. Specifically: Aim 2a. Healthcare utilization, measured by the need for emergency and acute care Aim 2b. Impact on cancer care delivery, specifically chemotherapy treatment duration and delays Aim 2c. Patients' outcomes, indicated by levels of self-efficacy and symptom burden Aim 2d. Patients' satisfaction with their cancer care
Aim 3: Evaluate the facilitators and barriers to implementation of an EHR-integrated ePRO symptom management system from the patient, clinician, and organizational perspectives. Specifically: Aim 3a. Patient adoption (including program feedback and experiences via qualitative interviews), clinician utilization, and their perspectives on appropriateness and acceptability Aim 3b. The sustainability of ePRO symptom management within a health system Aim 3c. Penetration and scalability of ePROs for symptom management Aim 3d. Extent of adaptation of ePRO systems over the course of the implementation process

2. Objectives of SW-RCT

This SW-CRT is conducted to determine the effectiveness of eSyM on health outcomes including healthcare utilization, measured by the need for emergency and acute care; impact on cancer care delivery, specifically chemotherapy treatment duration and delays; patients' outcomes, indicated by levels of self-efficacy and symptom burden; and patients' satisfaction with their cancer care.

3. Study Design

This study employs a type II hybrid effectiveness-implementation stepped wedge cluster randomized design. The research team will partner with software developers at Epic to adapt working ePRO symptom management systems, one in surgical and one in medical oncology, and fully integrate them into the EHR at 6 health systems. After pilot testing, we will conduct a pragmatic stepped wedge cluster randomized trial to measure the effectiveness of the ePRO system on outcomes that matter to patients and clinicians. Throughout, we will evaluate the implementation process to optimize sustainability and generate actionable knowledge that facilitates scaling to other settings. The proposed study is a hybrid

effectiveness-implementation type II trial according to the Curran schema.[1] The study meets PRECIS-2 criteria for pragmatic trials based on scores of 4 or higher in each domain. [2] Reports will adhere to the revised Criteria for Reporting the Development and Evaluation of Complex Interventions in healthcare (CReDECI2), [3] the CONSORT PRO[4] and cluster randomized extensions,[5] and the stepped wedge reporting guidelines proposed by Grayling.[6]

3.1. Randomization

This SW-RCT has 6 steps (figure below), and we randomize 6 sites (**Table 2**) to each of these steps. To ensure that two sites with the same region were not assigned to the same rollout step group (medical oncology live before surgery or surgery live before medical oncology), we employed a stratified randomization by region. As a result, each of the two groups has a site from each of the three regions (Northern, Southern, and Metropolitan). The stepped wedge design includes seven time-periods, including a run-in period (Figure).

Figure: SIMPRO Stepped-Wedge Randomization Schema

Abbreviations: BAPT=Baptist, WVU=West Virginia University, MMC=Maine Medical Center, DHMC=Dartmouth Hitchcock Medical Center, LCI=Lifespan Cancer Institute, DFCI=Dana-Farber Cancer Institute, SIMPRO=Symptom Management IMplementation of Patient Reported Outcomes in Oncology

Sequence	Group	Site	eSyM version	Period 1 Mar '19 - Aug '19	Period 2 Sep '19- Feb '20	Period 3 Mar '20- Aug '20	Period 4 Sep '20 – Feb '21	Period 5 Mar '21- Aug '21	Period 6 Sep '21- Feb '22	Period 7 Mar '22- Aug '22
Med Onc live Before Surgery	Southern	BAPT	Med	Pre-live	Live					
			Surg	Pre-live						Live
	Northern	MMC	Med	Pre-live		Live				
			Surg	Pre-live					Live	
	Metro-politan	DFCI	Med	Pre-live			Live			
			Surg	Pre-live				Live		
Surgery live Before Med Onc	Metro-politan	LCI	Med	Pre-live				Live		
			Surg	Pre-live			Live			
	Northern	DHMC	Med	Pre-live					Live	
			Surg	Pre-live		Live				
	Southern	WVU	Med	Pre-live						Live
			Surg	Pre-live	Live					

Note: The timing of medical and surgical rollouts is not the same at each site in this SW-RCT. The timing of medical and surgical rollouts was determined so that the number of subjects on the intervention and control would be the same in each site, when medical and surgical cohorts were combined.

Table 2: Six participating sites and Region

Site	Region
WVU	Southern
BAPT	
DHMC	Northern
MMC	
DFCI	Metropolitan
LCI	

3.2. Eligibility of SW-RCT participants

- Age ≥ 18 years
- Patients who meet one of the following:
    - Suspected thoracic cancer AND is inpatient following thoracic surgery.
    - Suspected gastrointestinal cancer AND is inpatient following gastrointestinal surgery.
    - Suspected gynecologic cancer AND is inpatient following gynecologic surgery.
    - Diagnosis of thoracic cancer AND scheduled to start a new treatment plan for thoracic cancer.
    - Diagnosis of gastrointestinal cancer AND scheduled to start a new treatment plan for gastrointestinal cancer.
    - Diagnosis of gynecologic cancer AND scheduled to start a new treatment plan for gynecologic cancer.

** Note 1: Patients undergoing thoracic, gynecologic, or gastrointestinal surgery may not be diagnosed with cancer. These patients are still eligible for eSyM usage, questionnaire completion, and medical record abstraction.*
** Note 2: Any patient at any participating site is allowed to be enrolled but those patients who did not meet the criteria listed above will be excluded from the primary analysis. Currently, only two sites are enrolling patients who do not meet the criteria listed above. If appropriate, we will perform the analysis with the data from all participants as exploratory analyses.*

3.3. Number of Subjects

Considering the 2016 tumor registry-reported analytic case volume at each study site for thoracic, GI and Gyn cancers, 12 patients per month per site for each of surgical and medical oncology is a highly conservative estimate of accrual. With 6 sites, assuming equal cluster size, 432 patients will be enrolled both for each period and for each cohort. Thus, a minimum of 6048 patients are expected to be enrolled to the SW-CRT. (**Table 3**) We will not have a cap for the number of participants for each site. Thus, the number of participants at the end of the study will be different from site to site.

Table 3: Conservative estimates of the accrual numbers for the SW-CRT

Site	MO/ Surg	Period 1 (Run-in)	Period 2	Period 3	Period 4	Period 5	Period 6	Period 7	**Total**
BAPT	MO	72	72	72	72	72	72	72	**504**
	Surg	72	72	72	72	72	72	72	**504**
MMC	MO	72	72	72	72	72	72	72	**504**
	Surg	72	72	72	72	72	72	72	**504**
DFCI	MO	72	72	72	72	72	72	72	**504**
	Surg	72	72	72	72	72	72	72	**504**
LCI	MO	72	72	72	72	72	72	72	**504**
	Surg	72	72	72	72	72	72	72	**504**
DHMC	MO	72	72	72	72	72	72	72	**504**
	Surg	72	72	72	72	72	72	72	**504**
WVU	MO	72	72	72	72	72	72	72	**504**
	Surg	72	72	72	72	72	72	72	**504**
Total	**MO**	**432**	**432**	**432**	**432**	**432**	**432**	**432**	**6048**
	Surg	**432**	**432**	**432**	**432**	**432**	**432**	**432**	

Notes:
- The length of each period is 6 months. The run-in period will be at least 6 months, and it will be adjusted so that the total number of eligible episodes under the control condition can be identical to that under the intervention condition.
- The cells colored by blue or grey indicate the periods where the intervention is rolled out in that hospital.

A subset of eSyM+ patients will be asked to complete a research questionnaire called the "SASS Questionnaire (eSyM+ version)" asking about their **S**elf-efficacy, **A**ttainment of information needs, **S**ymptom burden, and **S**atisfaction with care.

Table 4: SASS Questionnaire accrual numbers (Survey Cohort)

Site	Surgery			Medical Oncology			Totals
	eSyM- Version	eSyM+ Version	eSyM+ Non-Responder Version	eSyM- Version	eSyM+ Version	eSyM+ Non-Responder Version	
Site 1	75	75	15	75	75	15	330
Site 2	75	75	15	75	75	15	330
Site 3	75	75	15	75	75	15	330
Site 4	75	75	15	75	75	15	330
Site 5	75	75	15	75	75	15	330
Site 6	75	75	15	75	75	15	330
Totals	450	450	90	450	450	90	1,980
eSyM+ : Intervention condition eSyM- : Control condition							

The questionnaire will stop being administered once a minimum of 1,980 total surveys have been received in accordance with the above breakdown. (**Table 4**) ** *Total number of SASS participants through surveys can be larger or smaller depending on availability.*

A small subset of eSyM- and eSyM+ patients will be invited to take part in a one-time qualitative interview. Patients may or may not have previously completed SASS or eSyM questionnaires. Interviews will continue until thematic saturation is reached, or until 100 interviews are completed, whichever is reached first.

3.4. Duration of subject's participation in the study

1) Per protocol, eSyM usage continues for up to 60-180 days from the 1st trigger event but can continue indefinitely at the site's discretion.
2) The SASS Questionnaire is a one-time, 20-minute survey administered 30-180 days after surgery or first dose of chemotherapy.
3) The patient qualitative interview is a one-time, 30–60-minute interview administered any time after eSyM assignment.
4) The patients will be followed for outcomes for up to 1-year after chemotherapy starts (for MO) and discharge (for Surg).

3.5. Duration anticipated to enroll all study subjects

We anticipate that it will take four years to complete this activity.

3.6. Extension of accrual period

Due to COVID, the number of observed ER events was smaller than the expected number in three SIMPRO sites as of April 2022. The study team decided to extend accrual an extra 6 months (1 additional study period) to allow for additional event collection and patient recruitment. The extension will also allow sites who recently went live to continue program optimization and stabilization. The figure (below) depicts the revised study design with Period 8.

Sequence	Group	Site	eSyM version	Period 1 Mar '19 - Aug '19	Period 2 Sep '19- Feb '20	Period 3 Mar '20- Aug '20	Period 4 Sep '20 – Feb '21	Period 5 Mar '21- Aug '21	Period 6 Sep '21- Feb '22	Period 7 Mar '22- Aug '22	Period 8 Sept '22 – Feb '23
Med Onc live Before Surgery	Southern	BAPT	Med	Pre-live	Go-Live – 9/10/2019						
			Surg	Pre-live						Go-Live – 4/25/2022	
	Northern	MMC	Med	Pre-live		Go-Live – 3/16/2020					
			Surg	Pre-live					Go-Live – 10/19/2021		
	Metro-polytan	DFCI	Med	Pre-live			Go-Live – 9/22/2020 (2 clinics), 11/17/2020 (1 clinic)				
			Surg	Pre-live				Go-Live – 5/4/2021 (1 clinic), 6/15/2021 (2 clinics), 10/12/2021 (1 clinic)			
Surgery live Before Med Onc	Metro-polytan	LCI	Med	Pre-live				Go-Live – 9/21/2021			
			Surg	Pre-live			Go-Live – 11/24/2020				
	Northern	DHMC	Med	Pre-live					Go-Live – 11/30/2021		
			Surg	Pre-live		Go-Live – 4/28/2020					
	Southern	WVU	Med	Pre-live						Go-Live – 4/18/2022	
			Surg	Pre-live	Go-Live – 10/25/2019						

4. Analysis for Aim 2a: Healthcare utilization, measured by the need for emergency and acute care

4.1. Outcomes and summary measures

- **Table 5** shows the outcomes for Aim 2 and the summary measure to quantify the treatment effect magnitude for each outcome. The treatment effect magnitude on each outcome will be reported in both absolute and relative terms. terms (CONSORT Checklist Item 17b).[7]
- The primary outcome for Aim 2 is Emergency Department Treat/Release (EDTR) event occurrence status at 30-day post-chemotherapy start date (for MO) or at 30-day post-surgical discharge date (for Surg). The summary measure of this outcome is the proportion of the event occurrence at Day 30.
- Day 1 for each outcome is defined as the date of the initiation of chemotherapy for MO or the date of discharge from hospital (for Surg). However, events occurring on the first day of chemotherapy infusion or the day of discharge after surgery will be not counted as an "event," because they could not have been impacted by the eSyM program. By this definition, there will be no event at Day 1.

Table 5: Outcomes of Aim 2a

Outcome	Type of outcome	Summary measure for each group	Estimand (summary measure for between-group difference)	Cohort
Emergency Department Treat/Release (EDTR) event occurrence status at Day 30 [**primary**]	Binary	Proportion	Absolute Difference (AD) and Odds Ratio (OR)	MO and Surg
EDTR event occurrence status at Day 90	Binary	Proportion	AD and OR	MO and Surg
Number of the EDTR event occurrences during 180-day of follow-up	Count in a given follow-up period	Incidence rate on (0, 180d)	Incident Rate Difference (IRD) and Incidence Rate Ratio (IRR)	MO and Surg
Number of the EDTR event occurrences during 365-day of follow-up	Count in a given follow-up period	Incidence Rate on (0, 365d)	IRD and IRR	MO and Surg
ED-hospitalization event occurrence status at Day 30	Binary	Proportion	AD and OR	MO and Surg
ED-hospitalization event occurrence status at Day 90	Binary	Proportion	AD and OR	MO and Surg

Number of the ED-hospitalization event occurrences during 180-day of follow-up	Count in a given follow-up period	Incidence rate on (0, 180d)	IRD and IRR	MO and Surg
Number of the ED-hospitalization event occurrences during 365-day of follow-up	Count in a given follow-up period	Incidence Rate on (0, 365d)	IRD and IRR	MO and Surg
Occurrence of 1st chemotherapy discontinuation during one-year follow-up period	Time to event	Incidence Rate on (0, 365d)	IRD and IRR	All MO
Occurrence of 1st admission during one-year follow-up period	Time to event	Incidence Rate on (0, 365d)	IRD and IRR	All MO
Occurrence of 1st re-operation during one-year follow-up period	Time to event	Incidence Rate on (0, 365d)	IRD and IRR	All Surg
Occurrence of 1st re-admission during one-year follow-up period	Time to event	Incidence Rate on (0, 365d)	IRD and IRR	All Surg
Death during one-year follow-up period	Time to event	Incidence Rate on (0, 365d)	IRD and IRR	MO and Surg
Composite of EDTR, Hospitalization, and Death event occurrence status at Day 30	Binary	Proportion	Absolute Difference (AD) and Odds Ratio (OR)	MO and Surg
Death during 30-day post-chemotherapy start date (for MO) or at 30-day post-surgical discharge date (for Surg)	Time to event	Incidence Rate on (0, 30d)	IRD and IRR	MO and Surg

Notes:

- The Day 1 is defined the chemotherapy start date for MO and the discharge date for Surgery.
- All outcomes are defined in relation to the date of discharge from hospital (for Surg) or the initiation date of a new chemotherapy regimen (for MO).
- Note that events occurring on the first day of chemotherapy or the day of discharge after surgery will be excluded, because they could not have been impacted by the eSyM program.

4.2. General Analytic Approach

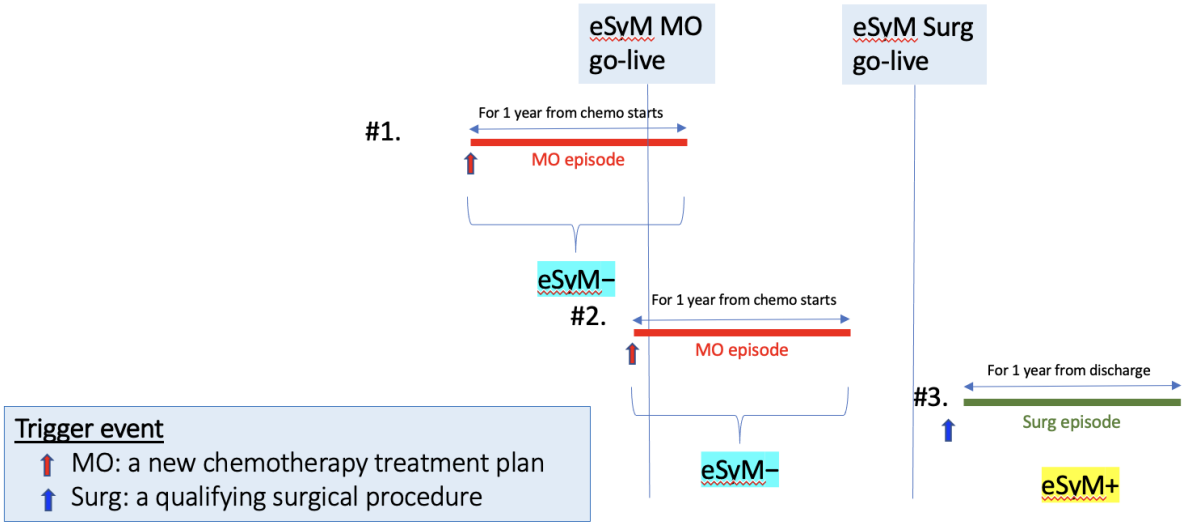
Medical Oncology and Surgery:

- The primary analysis will include data from both medical oncology (MO) and surgery (Surg) and estimate a common effect of the intervention in MO and SO. An indicator variable (MO vs. Surg), interaction between MO/Surg and Site, and interaction between MO/Surg and background time trend will be included as covariates for adjustment in multilevel generalized linear regression models.
- As a secondary analysis, for each outcome, the heterogeneity of the intervention effect by MO/Surg will also be assessed by the test for interaction between the intervention (eSyM yes/no) and MO/surg using multilevel generalized linear regression models. In this analysis, the interaction between the intervention and MO/Surg will be added to the model for the primary analysis.
- As another secondary analysis, we will analyze the data from MO and Surg separately and estimate the intervention effect for each outcome. Note that this analysis is similar to the aforementioned secondary analysis including the interaction between the intervention and MO/Surg. While the aforementioned secondary analysis assumes common effects of patients' characteristics (such as age, race, gender and so on) across MO and Surg, this analysis consider these are possibly different between MO and Surg.

Determination of the intervention and length of follow-up for each episode:

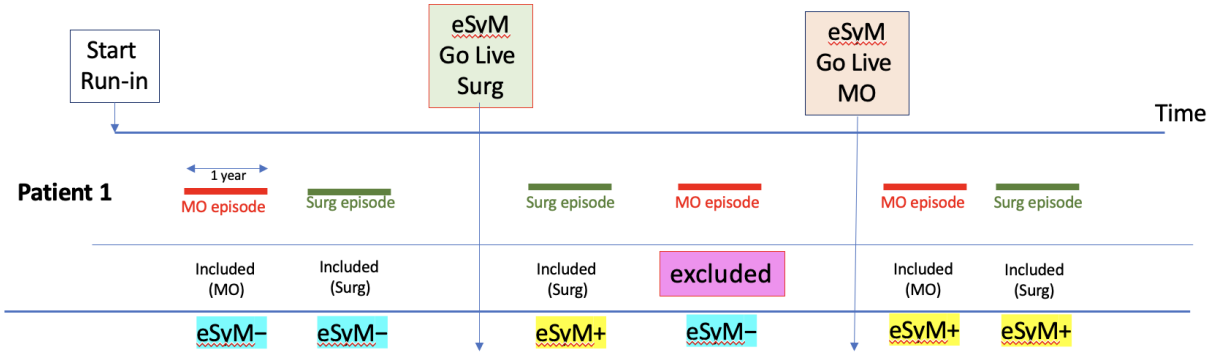
- The presence /absence of the intervention (eSyM+ vs. eSyM-) is determined in a cross-sectional way at the time of the occurrence of the trigger event
- The follow-up period for each episode is one year from the chemotherapy start date (**MO**) and the date of discharge (**Surg**)

- The intervention determined at the trigger event will never change during the one-year of follow-up period unless another trigger event occurs (see Figure below)



Multiple episodes for medical record data:

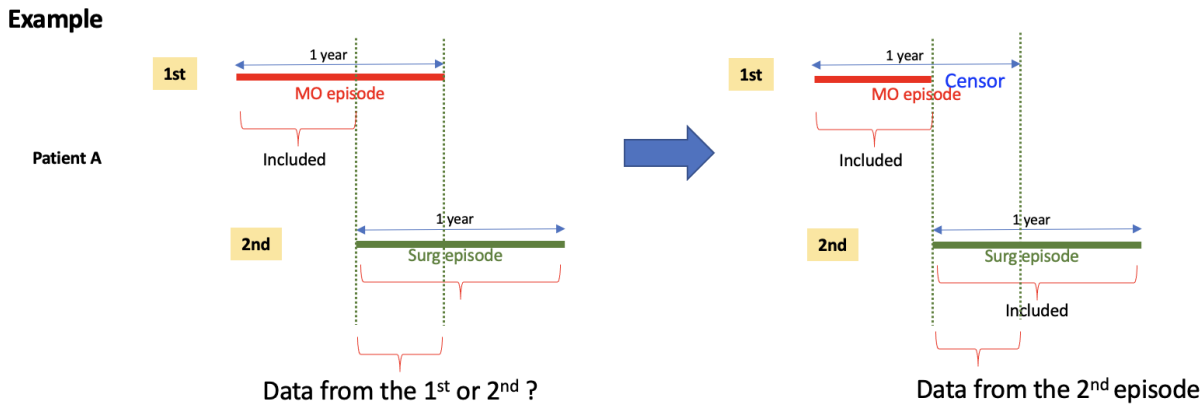
- Some subjects may experience more than one trigger event (i.e., a new chemotherapy treatment plan for **MO** and a qualifying procedure for **Surg**) during the study. Thus, potentially, multiple data points/episodes will be collected from one patient.
- The primary analysis will include all these episodes other than the following episodes:
  - MO episode where its trigger event occurred after the patient was exposed to eSyM Surg, but before eSyM MO was rolled out in the patient's site.
  - Surg episode where its trigger event occurred after the patient experienced eSyM MO, but before eSyM Surg was rolled out in the patient's site.
  - For example, the 4th episode in the Figure (below) was before go-live of MO in this site and thus this episode is considered eSyM control data(eSyM-) . While this patient did not experience the eSyM MO intervention at that time, this patient already experienced the eSyM Surg intervention (eSyM+) (see the 3rd episode in the figure). Therefore, we do not include the 4th episode as eSyM control (eSyM-) data but exclude it from the primary analysis.



- As sensitivity analyses, we will perform the analyses using the data consisting of only the first episode from each subject.

Occurrence of another trigger event during the follow-up of an episode:

- If another trigger event (Event B) occurs during the one-year follow-up period of a trigger event (Event A), the data collected after Event B will be included in the aggregation as data of the episode triggered by Event B, not by Event A (Figure below)



Analysis Populations:

All eligible subjects accrued in this study will be included in the primary analysis. To investigate the impact of patient's engagement with eSyM on clinical outcomes, two additional analysis populations will be considered for secondary and exploratory analyses.

To this end, we consider 3 total groups for analysis:

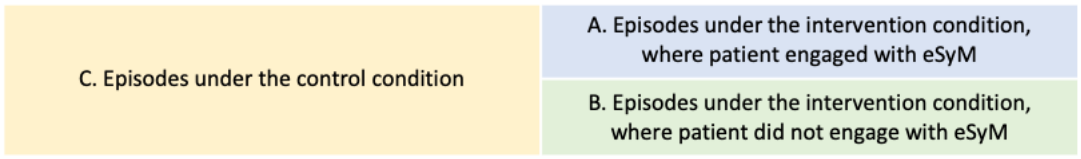
- Group A: episodes where subjects engaged* with eSyM among the eligible episodes of eligible subjects observed in the eSyM intervention (eSyM+) period.
- Group B: episodes where subjects did not engage with eSyM among the eligible episodes of eligible subjects observed in eSyM+.
- Group C: eligible episodes of eligible subjects observed in the eSyM control (eSyM-) period

* Definition of episodes where patient engaged with eSyM: episodes where patient reports ePRO through eSyM at least once during the follow-up

Table 6 shows the details of each analysis population.

Table 6: Analysis Populations and Objectives

Analysis Population	Comparison	Objective	Notes
1) A+B+C	(A+B) vs. C (Primary analysis)	Estimating the effect of eSyM exposure in a future population similar to the entire study population	Exposed to eSyM vs. Unexposed to eSyM regardless of engagement with the exposure (Effectiveness) Clinical outcome data from all groups will be included in the analysis on the basis of the Intention-to-treat principle.
2) A+C*	A vs. C (Pre-specified Secondary)	Estimating the effect of eSyM exposure in a future population that consists of patients who will engage with eSyM upon exposure.	Exposed to eSyM vs. Unexposed to eSyM conditional on engagement with eSyM (Efficacy) C* is a subset selected from C, so that the case-mix matches A (i.e., Engagement with eSyM). A and B are used to build the propensity score to select C*.
3) A+B	A vs. B (exploratory)	Estimating the effect of eSyM engagement in a future population similar to the study population if everyone engages, under the assumption that everyone in that population engages with eSyM upon exposure.	Engagement with eSyM vs. Unengagement with eSyM, conditioning on everyone being exposed to eSyM Propensity score weighting will be used to adjust for case-mix.



1) The primary analysis population consists of all eligible episodes under the intervention condition (A+B) (i.e., eSyM exposed) and all eligible episodes under the control condition (C) (i.e., eSyM unexposed).
2) The secondary analysis population consists of the episodes where patients engaged with eSyM (A) and the matched episodes during the pre-intervention period (C*). C* is a subset of C. In this analysis, we are interested in estimating the intervention effect among a subgroup of patients who will engage with eSyM if exposed. First, for MO and Surg, a propensity score model (a logistic regression model) to predict engagement with eSyM will be derived using the data from A and B. Next, for each subject in A and C, the predicted probability of engagement will be calculated using the derived propensity score model. The pre-specified variables used for the propensity score models are listed in **Table 7**. For each intervention episode in A, one control episode in C will then be matched, without replacement, based on their predicted probabilities of engagement, using the nearest neighbor matching method stratified by site. The matched dataset is then run through the same outcome model as the primary comparison.
3) As exploratory analyses, we will compare A vs. B. Under the condition that the hospitals have switched on the eSyM system (i.e., everybody is exposed to the eSyM system), we are interested in estimating the effect of the eSyM engagement on the outcomes under the assumption that everyone engages with eSyM, compared to the case where no one engaged with eSyM in spite of eSyM exposure. A propensity score weighting will be used to adjust for the case-mix between engagement with eSyM and non-engagmenet with eSyM. We will build a propensity score model for MO and Surg, separately, using the data from A and B. The pre-specified variables used for the propensity score models are listed in **Table 7**. No outcome model will be involved in this analysis.

Table 7: Pre-specified variables included in the propensity score models

Variable	2) A vs. C*	3) A vs. B	Note
Goal of Chemotherapy (MO with curative goal, MO with palliative goal, Others)	X	X	Only the models for MO
Calendar Time (Secular trends) [continuous with natural cubic spline]		X	
Site (Treating Facility) (DFCI, MMC, DHMC, LCI, BAPT, WVU)	X	X	
Age (year) (fractional polynomials)	X	X	
Sex (male vs. female)	X	X	
Race/Ethnicity (white/non-Hispanic, black/non-Hispanic, Hispanic, Other/Unknown)	X	X	
Employment (Employed, disabled, retired, other)	X	X	
Marital status (married vs. not married)	X	X	
Rurality based on zip code (large metro, small metro, suburban, rural)	X	X	
Insurance type (Medicare only, Medicare and Medicaid, Medicaid only, Private, Other)	X	X	
Cancer type (GI, Gyn, Thoracic)	X	X	
Socio-economic status (quintiles by the zip code-based poverty level)	X	X	
Co-morbidity (Charlson comorbidity index from encounter diagnoses in the 12 months before the first eligibility date) (0, 1, 2+)	X	X	

4.3. Description of the Study Participants and Characteristics of Eligible Episodes

Disposition of the study participants:

Following the CONSORT guideline,[6,8] a CONSORT diagram will be generated to describe the patient population with descriptions of the numbers of participants.

Patient demographic and disease characteristics:

- Baseline patient characteristics and disease characteristics at the start of follow-up will be summarized using descriptive statistics by sequence (i.e., site) .
- The unit of the analysis is not patient but episode. All eligible episodes eSyM- episodes will be included in this analysis.
- The same analysis will be performed for each of the two cohorts (MO and Surg).

The PRECIS score

The Pragmatic-Explanatory Continuum Indicator Summary (PRECIS) score will be calculated and reported.

4.4. Multilevel Generalized Linear Regression Analysis

The primary analysis for the outcomes in Aim 2 (Table 5) will be performed via multilevel generalized linear regression models (or generalized linear mixed-effects model; GLMM).[9] The null hypothesis for the statistical test is no intervention effect on the outcome, which indicates that the regression coefficient for the intervention indicator variable is zero. The intervention effect will be summarized in both absolute and relative terms (CONSORT Checklist Item 17b).[7]

Followings are specifications of the GLMM for the primary and secondary analyses.

- Link function
    - For binary outcomes, the logit-link will be used.
    - For count outcomes, the log-link will be used, and the follow-up time will be included as the offset in the model. (Poisson regression)
    - For time-to-event outcomes, we will use Poisson regression models, where the log-link will be used. The event indicator (1: event, 0: censoring) will be the response variable. Exposure time (i.e., days from the start date of the follow-up to either the date of the event occurrence or the end of follow-up, whichever comes first) will be included as the offset in the model.
- Treatment effect metrics
    - Following the recommendation of the CONSORT guideline, [7] we will report the treatment effect in both absolute and relative metrics.
    - For binary outcomes, we will report odds ratio (OR) derived from the GLMM with the logit-link. In addition, we will calculate absolute difference (AD), using the resulting GLMM with G-computation. Bootstrap will be used to obtain a standard error for the estimated AD.
    - For count outcomes and time-to-event outcomes, we will report the incidence rate ratio (IRR) derived from the GLMM model. In addition, we will calculate absolute difference in incidence rate (i.e., incidence rate difference; IRD) in a similar way to calculate AD.
- Secular trend
    - Primary: Calendar time of the start date of the follow-up of each episode will be included as a continuous variable with natural cubic spline with 3 knots
    - Secondary: Period in which the trigger event occurs will be included as a categorical variable.
- Site (cluster)

- Since there are only 6 sites in this study, the conventional analysis of including cluster as random-effects is problematic due to this small number.[10]
- Primary: The primary analysis includes sites as fixed effects.
- Secondary: Sites will be included as random-effects with a small sample adjustment (the Kenward-Roger method). [11,12]

- Subjects
  - We will potentially have more than one data point from one subject in this study.
  - To take account of within-subject correlation, subjects will be included as random effects.
  - The primary analysis will consider random intercepts only.
- Variables to be included the GLMM
  - **Table 8** lists the variables selected based on the study design (Category A) and clinical importance (Category B and C).
  - The primary analysis includes the variables in Category A and B.
  - As a secondary analysis, we will perform the same analysis including Category A, B and C.

Table 8: Variables that will or may be included in the generalized linear mixed-effects model for the primary analysis

Variable	Category A	Category B	Category C
Intervention indicator: (eSyM+ vs. eSyM-)	X		
Cohort (MO vs. Surgery)	X		
Calendar Time (Secular trends) [continuous with natural cubic spline]	X		
Site (Treating Facility) (DFCI, MMC, DHMC, LCI, BAPT, WVU) [as fixed effects]	X		
Interaction between Cohort (MO vs. Surgery) and Calendar Time (Secular trends)	X		
Interaction between Cohort (MO vs. Surgery) Site (Treating Facility) (DFCI, MMC, DHMC, LCI, BAPT, WVU) [as fixed effects]	X		
Subjects [as random effects (random intercepts only)]	X		
Goal of Chemotherapy (MO with curative goal, MO with palliative goal, Others)		X	
Age (years) (fractional polynomials)		X	
Sex (male vs. female)		X	
Race/Ethnicity (white/non-Hispanic, black/non-Hispanic, Hispanic, Other/Unknown)		X	
Rurality based on zip code (large metro, small metro, suburban, rural)		X	
Cancer/procedure type (GI, Gyn, Thoracic)		X	
Employment (Employed, Disabled, Retired, Other)			X
Marital status (Married vs. Not married)			X
Insurance type (Medicare only, Medicare and Medicaid, Medicaid only, Private, Other)			X
Socio-economic status (quintiles by the zip code-based poverty level)			X
Co-morbidity (Charlson comorbidity index from encounter diagnoses in the 12 months before the first eligibility date) (0, 1, 2+)			X

- Time decay of the intervention effect

- As a sensitivity analysis, we will include the interaction between the intervention and time from the rollout of the intervention condition at the treating site to the time of the trigger event, where we treat the time as a categorical variable in the model.

Considerations on the potential informative censoring

- Death is a competing event for EDTR and Hospitalization. In the primary analyses for EDTR and Admission, those who die without EDTR or Hospitalization will not be excluded from the analyses but included in the denominator when calculating the EDTR rate and Admission rate. Caution should be exercised in interpreting the intervention effect on these non-fatal outcomes. Because those who die before these non-fatal events will never experience it, the intervention effect on these non-fatal events will be biased in favor of either the control or intervention group where more deaths are observed. Therefore, when assessing the effectiveness of the intervention, we will also consider its impact on reducing death (a secondary outcome; see Table 5).
- As a sensitivity analysis for concluding the effectiveness of the intervention, we will analyze the composite of EDTR, Hospitalization, or Death event occurrence status at Day 30 (a secondary outcome; see Table 5), which is not subject to the competing risk problems.

Handling of potentially differentiable follow-up duration across episodes:

- While the planned follow-up duration for each episode is one-year, the actual follow-up durations of some episodes will be shorter than one year due to the occurrence of another trigger event or lost-to-follow-up during the one-year follow-up. Thus, the follow-up duration will be different across the episodes.
- For binary outcomes (i.e., the event status yes/no), the length of follow-up time will not be considered in the analysis. We expect that this will not affect the analysis of the primary endpoint (EDTR at Day 30) because we expect that we will have at least 30 days of follow-up for all episodes. The potential reason of censoring of follow-up is the occurrence of another trigger event during the follow-up time. It is unlikely for a patient to have another trigger event within 30 days after the previous trigger event.
- For count data, we will take the length of the follow-up period for each episode into account. Specifically, we will use Poisson regression models, where the logarithm of follow-up duration of each episode will be included as offset in the models.

4.5. Heterogeneity of effects and pre-specified subgroup analyses

With the small but clinically meaningful effect size we seek to detect, we have insufficient power to evaluate the EDTR rate in all subgroups of interest. Accordingly, subgroup analyses to assess for heterogeneity in treatment effects will be noted as exploratory and limited to pre-specified salient domains. Of greatest interest are:

1. Cohort (MO or Surg);
2. Region of Site (Northern, Southern, or Metropolitan)
3. Age [year] (<70 or >=70);
4. Sex (Male or Female);
5. Race (White/Non-Hispanic or Others);
6. Cancer type (GI, Gyn, or Thoracic); and
7. Rurality inferred from population density in the patient's zip code of residence (large metro, small metro, suburban, or rural).

4.6. Sensitivity analyses

Change in eSyM recall period:
During the study, there was a minor update to the eSyM symptom questionnaire. eSyM originally collected symptom information with a 7-day recall period, but a mid-study update altered this to a 24-hour recall period. Using eSyM+ data only, we will summarize frequencies of severe symptoms reported before the update and after the update. For the evaluation of eSyM intervention, we will not distinguish eSyM intervention before and after the update in the primary analysis. As sensitivity analyses, we will assess the impact of this update on each outcome.

- Sensitivity Analysis 1: We will run the same analysis as the primary analysis, excluding the eSyM+ episodes with a 7-day recall period.
- Sensitivity Analysis 2: We will modify the GLMM model used for the primary analysis. The intervention group variable will be a categorical variable with three levels – control, eSyM+ with 7-day recall, and eSyM+ with 24-hour recall. We will test if there is any difference between the 7-day and 24-hour recall.

Missing data:
Given the short 30-day interval for the Aim 2a primary endpoint and our pragmatic outcome, we expect the dropout fraction and missing data at Day 30 to be negligible. Subjects with no hospital encounters by Day 30 will be included in the analysis as having had no EDTR.

Regarding missing covariates, the primary analysis will be based on the following data handling.

1) Race/Ethnicity: Subjects with missing race/ethnicity are categorized as "Other/Unknown" and included in the analysis.
2) Employment: Subjects with missing employment information are categorized as "Other/Unknown" and included in the analysis.
3) Insurance type: Subjects with missing insurance type are categorized as "Other/Unknown" and included in the analysis.
4) Rurality: Subjects with missing rurality will be excluded from the analysis.
5) Socio-economic status: Subjects with missing socio-economic status are classified as "poverty level 2" and included in the analysis.

Note that the preliminary data check found that the other covariates do not have missing values.

In secondary analysis, we will perform multiple imputations with chained equations for Rurality. Specifically, the imputation models will include the primary outcome (EDTR Day 30), all variables used for the primary analysis (Category A and B), and the variables in C (as auxiliary variables) listed in Table 8. The number of complete datasets we generate will basically the same as the percentage of the fraction of missing. For example, if the missing fraction is 10%, we will generate 10 complete datasets. However, if the fraction of missing is less than 5%, we will generate 5 complete datasets. We will analyze each complete data and obtain the point estimate and the standard error for the intervention effects. We will then pool the results from the multiple complete datasets using Rubin's method.[13]

Analysis with data from the first episode:
In the primary analysis, all episodes in the analyses except for eSyM- episodes contaminated with the eSyM intervention will be included. We will perform sensitivity analyses using only the first episode from each patient. In these sensitivity analyses, each patient will serve only one data point in the analysis.

Modeling for the COVID19
As a sensitivity analysis, we will include the site-specific COVID19 prevalence as a time-varying covariate in the model.

4.7. Power Considerations

We estimate a minimum of 6048 patients will be enrolled. (see Section 3.3 for details). Since some subjects may experience more than one trigger event (see Section 4.2 General Analysis Approach: *Multiple episodes for medical record data),* the study participants will potentially contribute multiple data points. The goal of the power calculation for this study is to confirm that the study has sufficient power to assess the effects of the intervention on the outcomes. Therefore, for the power calculations throughout this SAP, we will conservatively estimate the power, assuming that an individual patient provides only one episode.

The probability that a patient who has a trigger event experiences EDTR within 30 days after the trigger event is estimated to vary between 8% to 15% for the control group, based on the HCUP data, institutional data, and early phase analyses from CMMI's Oncology Care Model for Baptist Memorial, the only Oncology Care Model participant among our 6 sites. We hypothesize that the probability of experiencing EDTR by Day 30 will be 3 to 4% lower in the eSyM+ group.

Table 9 shows the required sample sizes for the study to have 80% power to detect the difference between groups, at two-sided alpha level 0.05, using the SW-RCT design.[14] Analyses of inter-institutional variation in hospitalization rates and ED visit rates from AHRQ's statewide databases indicate that low ICC estimates are appropriate.[15,16] Furthermore, we expect a low ICC because: 1) the intervention will be deployed using the same technology across sites; 2) we will adjust for variation in baseline risk via GLMMs such that any potential differences in case-mix among sites will be negligible. Given these factors, the conservative estimate of 6048 participants provides adequate power to address the Aim 2a primary outcome.

Note that we used the sample size formula[14] that assumes exchangeable correlations because we originally had planned to include sites as random effects. However, since the number of sites is only 6 in this study, we modified the analysis plan and decided to include sites as fixed effects in the primary analysis. We acknowledge that our original sample size calculation (above) may not be precise due to this, but it should still work as a conservative estimate.

Table 9. Required total sample size with various scenarios for 80% power

The probability of experiencing EDTR by Day 30		Effect size		N required	
Control (eSyM-)	Intervention (eSyM+)	Absolute	Relative	Low ICC (0.01)	Moderate ICC (0.05)
8%	5%	3%	38%	2192	5844
9%	6%	3%	33%	4485	6816
10%	7%	3%	30%	5047	7721
11%	8%	3%	27%	5598	8606
12%	8%	4%	33%	3246	4810
13%	9%	4%	31%	3538	5283
14%	10%	4%	29%	3823	5746
15%	11%	4%	27%	4103	6198

4.8. Table shell for presenting the primary analysis results

MO and Surgery Combined						
	Number of Observed Events (%)		Odds Ratio		Absolute Difference	
Outcomes	Intervention (#patients=XX) (#episodes = XX)	Control (#patients=XX) (#episodes = XX)	Estimate (0.95 CI)	p-value	Estimate (0.95 CI)	p-value
EDTR Day 30	XX (XX%)	XX (XX%)	X.XX (X.XX to X.XX)	X.XXX	X.XX (X.XX to X.XX)	X.XXX
EDTR Day 90	XX (XX%)	XX (XX%)	X.XX (X.XX to X.XX)	X.XXX	X.XX (X.XX to X.XX)	X.XXX
Admission Day 30	XX (XX%)	XX (XX%)	X.XX (X.XX to X.XX)	X.XXX	X.XX (X.XX to X.XX)	X.XXX
Admission Day 90	XX (XX%)	XX (XX%)	X.XX (X.XX to X.XX)	X.XXX	X.XX (X.XX to X.XX)	X.XXX
MO Only						
	Number of Observed Events (%)		Odds Ratio		Absolute Difference	
Outcomes	Intervention (#patients=XX) (#episodes = XX)	Control (#patients=XX) (#episodes = XX)	Estimate (0.95 CI)	p-value	Estimate (0.95 CI)	p-value
EDTR Day 30	XX (XX%)	XX (XX%)	X.XX (X.XX to X.XX)	X.XXX	X.XX (X.XX to X.XX)	X.XXX
EDTR Day 90	XX (XX%)	XX (XX%)	X.XX (X.XX to X.XX)	X.XXX	X.XX (X.XX to X.XX)	X.XXX
Admission Day 30	XX (XX%)	XX (XX%)	X.XX (X.XX to X.XX)	X.XXX	X.XX (X.XX to X.XX)	X.XXX
Admission Day 90	XX (XX%)	XX (XX%)	X.XX (X.XX to X.XX)	X.XXX	X.XX (X.XX to X.XX)	X.XXX
Surgery Only						
	Number of Observed Events (%)		Odds Ratio		Absolute Difference	
Outcomes	Intervention (#patients = XX) (#episodes = XX)	Control (#patients=XX) (#episodes = XX)	Estimate (0.95 CI)	p-value	Estimate (0.95 CI)	p-value
EDTR Day 30	XX (XX%)	XX (XX%)	X.XX (X.XX to X.XX)	X.XXX	X.XX (X.XX to X.XX)	X.XXX
EDTR Day 90	XX (XX%)	XX (XX%)	X.XX (X.XX to X.XX)	X.XXX	X.XX (X.XX to X.XX)	X.XXX
Admission Day 30	XX (XX%)	XX (XX%)	X.XX (X.XX to X.XX)	X.XXX	X.XX (X.XX to X.XX)	X.XXX
Admission Day 90	XX (XX%)	XX (XX%)	X.XX (X.XX to X.XX)	X.XXX	X.XX (X.XX to X.XX)	X.XXX
Factors used for the adjustment: XXX, XXX, XXX, XXX						

5. Analysis for Aim 2b: Impact on initiation of adjuvant chemotherapy and chemotherapy duration.

5.1. Outcomes and summary measures

We do not have preliminary data to support a specific effect size. However, we expect that patients exposed to eSyM may be able to: 1) initiate adjuvant therapy sooner; and/or 2) remain on their chemotherapy regimens for longer duration. These time intervals are straightforward to measure from EHR encounter and date fields. The outcomes for Aim 2b and the projected minimum sample size are summarized in **Table 10**.

Table 10: Outcomes of Aim 2b

Outcome	Type of outcome	Summary measure for each group	Summary measure for between-group difference	**Cohort**	**Projected *minimum* sample size**
Days from chemo regimen start date to stop date	Continuous	Mean*	Absolute Difference (AD)	All MO	504 patients/site; 3024 MO patients in total 1512 eSyM+ and 1512 eSyM-
Days from surgery date to start of adjuvant chemotherapy	Continuous	Mean*	AD	Surg who get adjuvant Rx (roughly 40% of all Surg)	202 patients/site; 1212 surg patients in total 606 eSyM+ and 606 eSyM-

* Because the maximum follow-up period of each episode is 1 year, the data will be truncated at 365 days when the mean value is calculated.

For MO patients, the outcome is time from the first dose to the last dose of a specific regimen. We will censor follow-up at 1 year. For Surg, the denominator population consists of patients who receive *any* adjuvant chemotherapy within 6 postoperative months. Tumor registry stage distribution at our 6 sites indicates that this will be 202 patients per site or 1212 in total.
The type of censoring involved in this analysis is only the truncation at the maximum follow-up period of 1 year. Thus, we will not employ censored time-to-event analysis methods but handle the outcomes as continuous outcomes.

5.2. Data Analysis

Similar analyses to those for Aim 2a (see Section 4) will be performed. GLMMs will be used with the indemnity link function for these continuous variables. The intervention effect will be summarized as the mean difference between eSyM+ and eSyM- and reported with a corresponding 95% confidence interval and p-value.

5.3. Power Considerations

Our general approach is to determine the magnitude of difference detectable at a given sample size. Because we have insufficient preliminary information about the magnitude of the ICCs or standard deviations (SD) for our outcomes, we will consider a broad range.

For MO, assuming the average duration of chemotherapy is 180 days in the eSyM- group, the range of detectable between group differences is shown in **Table 11A**. Based on these very

conservative estimates, we have more than 80% power to detect a 16% increase in chemotherapy duration, which corresponds to nearly 1 month.

For Surg, we assume that the average duration from surgery to adjuvant chemotherapy start date (censored at 6 months) is 56 days (8 weeks) in the eSyM- group. With 1212 patients (see **Table 10**), and conservative parameter estimates, we have >80% power to detect a 25% increase in the interval between surgery and starting adjuvant therapy. This corresponds to 14 days, a difference that is both plausible and meaningful, see **Table 11B.**

Table 11: Power considerations for the outcomes of Aim 2b

A: MO 3024 patients (1512 per group)

Detectable effect size with 80% power	Assumed ICC	Assumed standard deviation [Day]	Difference in time to discontinuation	Corresponding ratio in the eSyM- group
0.138	0.01% (Very low)	202 (conservative)	28 days	1.16
		152	21 days	1.12
		101	14 days	1.08
0.170	1%	165 (conservative)	28 days	1.16
		124	21 days	1.12
		82	14 days	1.08
0.174	5% (conservative)	161 (conservative)	28 days	1.16
		120	21 days	1.12
		80	14 days	1.08

B: Surg 1212 patients (606 per group)

Detectable effect size with 80% power	Assumed ICC	Assumed standard deviation	Difference in time to adjuvant chemo	Corresponding ratio in eSyM- group
0.217	0.01% (Very low)	65 (conservative)	14 days	1.25
		32	7 days	1.13
0.256	1%	55 (conservative)	14 days	1.25
		27	7 days	1.13
0.271	5% (conservative)	52 (conservative)	14 days	1.25
		26	7 days	1.13

6. Analysis for Aim 2c: Patients' outcomes, indicated by levels of self-efficacy and symptom burden

6.1. Outcomes and summary measures

The outcomes investigated in Aim 2c of the SIMPRO study are classified into four groups, depending on the data source and data collection schedule. **Table 12** shows the details of the groups and the disposition of each outcome in Aim 2c.

Table 12: Outcomes in Aim 2c

Type	Outcome Type in the SIMPRO study	Example outcomes in Aim 2c
1	Measured at one single time point only either in post-intervention (the intervention condition) or in pre-intervention (the control condition) for all or selected subjects (patients or providers)	eSyM decline (1st time)
2	Measured at multiple time points (periodically) only in post-intervention period for all or selected subjects (patients or providers)	eSyM logins per patient per month
		Frequency of patient eSyM reports with severe symptoms
3	Measured from selected subjects (patients or providers) at a single time point, which could be during the pre-intervention period or the post-intervention period. For some subjects, data may be collected at both periods.	PROMIS items (self-efficacy for managing symptoms, pain interference, fatigue, and physical function)
4	Measured periodically through the SW-CRT design from all eligible subjects.	MyChart activation (binary outcome)
		MyChart logins per patient per month

6.2. Data Analysis

Analysis of Outcomes classified as Type 1:

- For descriptive purposes, we will estimate crude means for continuous outcomes and crude proportions for dichotomous outcomes and calculate corresponding 95% confidence intervals by site. The coefficient of variation is calculated using the site-specific point estimates.
- Next, we estimate the overall mean (or proportion). To this end, a weighted mean and corresponding 95% confidence interval will be calculated using the site-specific point estimates and their standard error estimates, using the inverse of the site-specific variance estimates for the weights.
- In addition, generalized linear models (GLMs) will be used to investigate factors associated with the outcome. In the GLMs, we will account for site variability by treating "site" as a fixed effect. For modeling continuous outcomes, we will use the identity link function. For dichotomous outcomes, the logit link function is used.

Analysis of Outcomes classified as Type 2:

- We will calculate the site-period-specific means (or proportions), adjusted for the factors potentially associated with the outcome. We will calculate the coefficient of variations across sites and also across periods, respectively.
- To this end, we will use generalized linear mixed-effects models (GLMMs). In GLMMs, we will account for site variability by treating "site" as a fixed-effect, secular trend by treating "period" as a fixed-effect, and within-subject variability by treating "subject" as random-effects. For modeling continuous outcomes, we will employ the identity link function. For

dichotomous outcomes, the logit link function will be utilized. G-computation will be used to derive the site-period-specific means (or proportions).

- We will also calculate the site-specific, period-specific, and overall means (or proportions) and corresponding 0.95 confidence interval by taking a weighted average of the site-period-specific mean values (or proportions) obtained through GLMMs. The reciprocal of the variance-covariance matrices of the site-period-specific mean values (or proportions) will be used as the weight.

Analysis of Outcomes classified as Type 3:

- For descriptive purposes, we will estimate crude means for continuous outcomes and proportions for dichotomous outcomes and calculate corresponding 95% confidence intervals by intervention group (eSyM+ vs. eSyM-) and site.
- Using these results, the difference between the intervention condition (eSyM+) and the control condition (eSyM-) along with a corresponding 95% confidence interval will be calculated for each site. In addition, using the point estimates from the six sites, we will calculate the coefficient of variation of the difference between conditions.
- We will then estimate the overall between-conditions difference integrating the site-specific differences. Specifically, a weighted average and corresponding 95% confidence interval will be calculated, using the site-specific point estimates their standard error estimates, where the reciprocal of the site-specific variance estimates will be used for the weights.
- In addition, generalized linear models (GLMs) will be used to investigate factors associated with the outcome. In the GLMs, we will include the intervention indicator, "site" and other factors (i.e., age (at cancer diagnosis for medical oncology patients and at index surgery for surgical patients), sex, employment status, education, ability to pay bills, and technology confidence) as independent variables. For modeling continuous outcomes, we will employ the identity link function. For dichotomous outcomes, the logit link function will be utilized. Note that, when more appropriate (where some subjects have data from both pre- and post-intervention periods), we will use GLMMs by treating "subject" as random-effects instead of GLMs for some outcomes.

Analysis of Outcomes classified as Type 4:

- We will estimate the intervention effect on each outcome.
- To this end, analyses similar to those for Aim 2a (see Section 4) will be conducted. GLMMs will be used with the identity link function for continuous outcomes and with the logit link function for dichotomous outcomes. The intervention effect will be summarized as the mean difference and odds ratio for continuous outcomes and dichotomous outcomes, respectively, and reported with a corresponding 95% confidence interval and p-value.

Interpretation of the PROMIS scores: Note that Yost and Cella have reported minimally important difference (MID) ranges for five PROMIS domains including fatigue, pain, depression, anxiety, and physical functioning[17,18] Cella recommends using 0.5 SD as the MID for PROMIS scales[19,20]

Approach to missing data for PROMIS scores: Because a random missing mechanism assumption is not verifiable, we will use several methods to handle missing observations. Specifically, we will perform: (1) mean value; (2) worst-case; (3) best-case; and, (4) multiple imputations.[13]

7. Analysis for Aim 2d: Patients' satisfaction with their cancer care

7.1. Outcomes and summary measures

The questionnaire asks patients to report on satisfaction with their cancer care using the CAHPS Cancer Care Survey. This outcome is classified as the Type 3 outcome in Table 12.

7.2. Data Analysis

Aim 2d survey methods and sampling mirrors Aim 2c (PROMIS items). Following Section 6.2 (Analysis of Outcomes classified as Type 3), we will perform the analysis for these continuous outcomes. We will use the AHRQ's CAHPS Analysis Program.[21,22]

8. Analysis for Aim 3a: Patient adoption, clinician utilization, and their perspectives on appropriateness and acceptability

8.1. Outcomes and summary measures

The outcomes investigated in Aim 3a in the SIMPRO study are classified into four groups, depending on the data source and data collection schedule. **Table 13** shows the details of the groups and the disposition of each outcome in Aim 3a.

Table 13: Outcomes in Aim 3a

Type	Outcome Type in the SIMPRO study	Example outcomes in Aim 3a
1	Measured at one single time point only either in post-intervention (the intervention condition) or in pre-intervention (the control condition) for all or selected subjects (patients or providers)	Qualitative feedback
		Pt satisfication with eSyM tool
		System Usability Items
		FIM item
		AIM items (acceptability)
		NOMAD items
		CSAT items
		# of med onc clinics continuing to use eSyM after stepped wedge period
		# of surg clinics continuing to use eSyM after stepped wedge period
		# of clinics that have >= 1 team conducting eSyM severe symptom outreach (e.g. using IB alerts)
		# of clinics that have >= 1 team conducting eSyM population symptom mgmt (e.g. using reports)
		Clicker questions (readiness + appropriateness for intervention) **[pre-intervention only]**
2	Measured at multiple time points (periodically) only in post-intervention period for all or selected subjects (patients or providers)	% of patients eligible to use eSyM completing 1+ qnr (PRIMARY IMPLEMENTATION OUTCOME)
		# pts assigned to eSyM/total # of pts with index condition on eSyM registry
		# reporting eSyM once, 25%, 50%, 75% of prompts
		Clinical staff responses to eSyM reports (InBasket + Telephone Encounters + Messages)
		Patient eSyM Total Usage Rate (w/MyChart) - up to 1 year after go-live
		Patient eSyM Total Usage Rate (w/ or w/o MyChart) - up to 1 year after go-live
		Patient eSyM Weekly Usage Rate (w/MyChart) - up to 1 year after go-live
		Patient eSyM Weekly Usage Rate (w/ or w/o MyChart) - up to 1 year after go-live
		% of MyChart Patients w/ documented outreach - up to 1 year after go-live
		% of patients with eSyM assigned w/ documented outreach - up to 1 year after go-live
		% of weekly patients responding to eSyM w/ documented outreach - up to 1 year after go-live
		% of weekly responders reporting moderate-severe symptoms - up to 1 year after go-live
		% of weekly responders reporting severe symptoms - up to 1 year after go-live

3	Measured from selected subjects (patients or providers) at a single time point, which could be during the pre-intervention period or the post-intervention period. For some subjects, data may be collected at both periods.	program barriers (stakeholder survery)
		Program facilitators (stakeholder survery)
4	Measured periodically through the SW-CRT design from all eligible subjects.	# of telephone encounter per patient per month

Notes:

- Both patient adoption and clinician utilization can be observed from analyzing EHR data based on eSyM utilization patterns. Clinician utilization can also be measured from the EHR and will be grouped in categories.

- Data for Aim 3a (clinicians) will be collected from the EHR based on system utilization patterns as well as from qualitative and quantitative survey, especially interviews with participating staff and clinicians.

- Appropriateness and acceptability will be ascertained using Weiner's AIM surveys (8-items total) which will be administered along with CAHPS surveys. Appropriateness and acceptability ratings will be defined based on the % of respondents who "agree" or "completely agree" with the survey items compared to the % who are neutral, disagree, or completely disagree and characterized using descriptive statistics.[23]

8.2. Data Analysis

Analysis of Outcomes classified as Type 1: See Section 6.2

Analysis of Outcomes classified as Type 2: See Section 6.2

Analysis of Outcomes classified as Type 3: See Section 6.2

Analysis of Outcomes classified as Type 4: See Section 6.2

9. Analysis for Aim 3b: The sustainability of ePRO symptom management within a health system

Hypotheses:
(1) We hypothesize that 3 or more of our health systems will continue eSyM reporting for MO patients beyond 90 days.
(2) We hypothesize that tapering the dedicated nursing support provided by the study does not affect the effect of eSyM on clinical/utilization outcomes.

Analysis:

We will evaluate sustainability at the patient, clinic and health system level using simple rates and proportions. To evaluate sustainability, we will examine the consequences of withdrawing grant-funded nursing support for symptom management in the post-implementation period. We will compare outcomes from Period 6 (study month 45-50, all sites eSyM+) and the post-Implementation (Post-I; study months 51-56). Sites are trained and empowered to manage eSyM autonomously without research study staff. Then, during post-implementation, dedicated nursing support to monitor eSyM is *tapered* in half the sites (see Figure C2). To examine whether backing off on the study support attenuates the effect, we will perform difference in difference analysis. For each outcome, we will calculate the difference between Post-I and Period 6 outcomes by site. We then calculate the difference between site groups and the corresponding 95% confidence intervals.

10. Analysis for Aim 3c (Penetration and scalability of ePROs for symptom management) and Aim 3d (Extent of adaptation of ePRO systems over the course of the implementation process)

A mixed-methods approach:
Our analytic approach will integrate qualitative and quantitative data to obtain an informative description of factors that influence implementation at the level of 1) the practice site, 2) the health system, and 3) the entire project. The CFIR schema will facilitate comparisons, across settings and timepoints. First, we will analyze the qualitative and quantitative data separately. Next, we will use a weaving approach to integrate each source narrative to report findings on a theme-by-theme basis.[24]

Qualitative:
Analysis of interviews and group discussions will rely on notes and transcribed audiotapes entered in NVivo software. Given the study purpose, we will use a framework analysis[25–27] approach that allows for systematic analysis that is also flexible and iterative in nature.[28] Through indexing, charting, and mapping we will be able to draw comparisons across interviews to identify facilitators and barriers. Lastly, we will use the CFIR qualitative data scoring schema **(Table 14)** to assign a numeric code ranging from -2 to +2 to each construct summarizing whether it was a negative or positive influence on eSyM use. The weighted kappa statistic, with 0.70 as the cut-off will be used to assess agreement between coders. Ratings will be used to make topographic maps to visually convey the relative importance of each CFIR construct.[29]

Table 14: Rating the Influence of CFIR Constructs on Implementation Outcomes	
-2	Negative influence, impeding influence in implementation efforts
-1	Negative influence, impeding influence in implementation efforts (general impression but no concrete examples given during interviews; mixed effect)
0	Neutral influence; contradictory interviews
+1	Positive influence, facilitating influence in implementation efforts (general statements but no concrete examples; mixed effect, but generally positive)
+2	Positive influence, facilitating influence (firm examples shown or give)
missing	Lack of interviewee input or absence of evaluable construct

Quantitative:
Clinician surveys will use ordinal response scales to capture the valence (+/- influence) for each CFIR construct and simple descriptive statistics to characterize responses. Finally, variables organized by CFIR domain will be combined with qualitative data on a theme-by-theme basis. This will provide an interpretable numeric summary of the perceived importance of each construct. We will use 95%

confidence intervals as a measure of precision. All analyses at the health system level will adjust for clustering using generalized linear mixed effects models.

11. References

1. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care. 2012 Mar;50(3):217–226. PMCID: PMC3731143

2. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ. BMJ; 2015 May 8;350(may08 1):h2147. PMID: 25956159

3. Möhler R, Köpke S, Meyer G. Criteria for Reporting the Development and Evaluation of Complex Interventions in healthcare: revised guideline (CReDECI 2). Trials. Springer Science and Business Media LLC; 2015 May 3;16(1):204. PMCID: PMC4461976

4. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, for the CONSORT PRO Group. Reporting of Patient-Reported Outcomes in Randomized Trials [Internet]. JAMA. 2013. p. 814. Available from: http://dx.doi.org/10.1001/jama.2013.879

5. Campbell MK, Piaggio G, Elbourne DR, Altman DG, for the CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ. British Medical Journal Publishing Group; 2012 Sep 4;345(sep04 1):e5661–e5661.

6. Grayling MJ, Wason JMS, Mander AP. Stepped wedge cluster randomized controlled trial designs: a review of reporting quality and design features. Trials. 2017 Jan 21;18(1):33. PMCID: PMC5251280

7. Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, Dixon-Woods M, Aldcroft A, Doussau A, Grayling M, Kristunas C, Goldstein CE, Campbell MK, Girling A, Eldridge S, Campbell MJ, Lilford RJ, Weijer C, Forbes AB, Grimshaw JM. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. BMJ. 2018 Nov 9;363:k1614. PMCID: PMC6225589

8. Hemming K, Taljaard M, Grimshaw J. Introducing the new CONSORT extension for stepped-wedge cluster randomised trials [Internet]. Trials. 2019. Available from: http://dx.doi.org/10.1186/s13063-018-3116-3

9. Ariti C. Walter W Stroup, Generalized linear mixed models, modern concepts, methods and applications. Stroup Walter W , Generalized linear mixed models, modern concepts, methods and applications . CRC Press: Boca Raton, 2012; £59.99 p. 555. Stat Methods Med Res. SAGE Publications; 2017 Apr;26(2):1043–1044. PMID: 25549967

10. Elff M, Heisig JP, Schaeffer M, Shikano S. Multilevel analysis with few clusters: Improving likelihood-based methods to provide unbiased estimates and accurate inference. Br J Polit Sci. Cambridge University Press (CUP); 2021 Jan;51(1):412–426.

11. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics. 1997 Sep;53(3):983–997. PMID: 9333350

12. Li P, Redden DT. Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials. BMC Med Res Methodol. Springer Nature; 2015 Apr 23;15(1):38. PMCID: PMC4458010

13. Schafer JL. Multiple imputation: a primer. Stat Methods Med Res. SAGE PublicationsSage UK: London, England; 1999 Mar 1;8(1):3–15.

14. Hemming K, Taljaard M. Sample size calculations for stepped wedge and cluster randomised trials: a unified approach. J Clin Epidemiol. Elsevier Inc; 2016 Jan 1;69:137–146.

15. Campbell MK, Mollison J, Grimshaw JM. Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. Stat Med. 2001 Feb 15;20(3):391–399. PMID: 11180309

16. Eldridge SM, Costelloe CE, Kahan BC, Lancaster GA, Kerry SM. How big should the pilot study for my cluster randomised trial be? Stat Methods Med Res. SAGE Publications Ltd STM; 2016 Jun 1;25(3):1039–1056.

17. Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients [Internet]. Journal of Clinical Epidemiology. 2011. p. 507–516. Available from: http://dx.doi.org/10.1016/j.jclinepi.2010.11.018

18. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci. 2013 Oct 2;8:117. PMCID: PMC3852739

19. Temel JS, Greer JA, El-Jawahri A, Pirl WF, Park ER, Jackson VA, Back AL, Kamdar M, Jacobsen J, Chittenden EH, Rinaldi SP, Gallagher ER, Eusebio JR, Li Z, Muzikansky A, Ryan DP. Effects of Early Integrated Palliative Care in Patients With Lung and GI Cancer: A Randomized Clinical Trial. J Clin Oncol. 2017 Mar 10;35(8):834–841. PMCID: PMC5455686

20. Chen CX, Kroenke K, Stump TE, Kean J, Carpenter JS, Krebs EE, Bair MJ, Damush TM, Monahan PO. Estimating minimally important differences for the PROMIS pain interference scales: results from 3 randomized clinical trials. Pain. 2018 Apr;159(4):775–782. PMCID: PMC5860950

21. Hays R, Quigley D, Mendel P, Predmore Z, Chen AY. Use of CAHPS® patient experience survey data as part of a patient-centered medical home quality improvement initiative [Internet]. Journal of Healthcare Leadership. 2015. p. 41. Available from: http://dx.doi.org/10.2147/jhl.s69963

22. Quigley D, Qureshi N, Rybowski L, Shaller D, Edgman-Levitan S, Cleary PD, Ginsberg C, Hays RD. Summary of the 2020 AHRQ research meeting on 'advancing methods of implementing and evaluating patient experience improvement using consumer assessment of healthcare providers and systems (CAHPS®) surveys' [Internet]. Expert Review of Pharmacoeconomics & Outcomes Research. 2022. p. 883–890. Available from: http://dx.doi.org/10.1080/14737167.2022.2064848

23. Basch E, Snyder C. Overcoming barriers to integrating patient-reported outcomes in clinical practice and electronic health records. Annals of oncology: official journal of the European Society for Medical Oncology / ESMO. 2017. p. 2332–2333. PMID: 28961852

24. Snyder CF, Herman JM, White SM, Luber BS, Blackford AL, Carducci MA, Wu AW. When using patient-reported outcomes in clinical practice, the measure matters: a randomized controlled trial. J Oncol Pract. 2014 Sep;10(5):e299-306. PMCID: PMC4161731

25. Roess A. The Promise, Growth, and Reality of Mobile Health - Another Data-free Zone. N Engl J Med. 2017 Nov 23;377(21):2010–2011. PMID: 29116869

26. Jensen RE, Snyder CF. PRO-cision Medicine: Personalizing Patient Care Using Patient-Reported Outcomes. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2016. p. 527–529. PMID: 26644538

27. Hughes EF, Wu AW, Carducci MA, Snyder CF. What can I do? Recommendations for responding to issues identified by patient-reported outcomes assessments used in clinical practice. J Support Oncol. 2012 May 18;10(4):143–148. PMCID: PMC3384764

28. Cortez NG, Cohen IG, Kesselheim AS. FDA regulation of mobile health technologies. N Engl J Med. 2014 Jul 24;371(4):372–379. PMID: 25054722

29. Brundage M, Blackford A, Tolbert E, Smith K, Bantug E, Snyder C, PRO Data Presentation Stakeholder Advisory Board (various names and locations). Presenting comparative study PRO results to clinicians and researchers: beyond the eye of the beholder. Qual Life Res. 2018 Jan;27(1):75–90. PMCID: PMC5770492

ADDENDUM TO THE STATISTICAL ANALYSIS PLAN

for

PROTOCOL ACTIVITY 4

Pragmatic stepped-wedge cluster randomized trial

SIMPRO Research Center: Integration and Implementation of PROs for Symptom Management in Oncology Practice.

DATE: *June 12, 2024*

Sponsor and Coordinating Center:

Dana-Farber/Harvard Cancer Center (DF/HCC)
Department of Medical Oncology
Coordinating Center PI and Technology PI: Michael Hassett
Dana-Farber Cancer Institute
Boston, MA 02215
Email Address: Michael_Hassett@dfci.harvard.edu
Telephone Number: 617-632-4587

CO-STUDY CHAIRS:

Deborah Schrag MD, MPH
Chair, Department of Medicine
Memorial Sloan Kettering Cancer Center
New York City, NY 10065

Raymond Osarogiagbon MD
Director, Thoracic Oncology Research Group
Baptist Memorial Hospital
Covington, Tennessee 38019

Sandra Wong MD
Chair, Department of Surgery
Dartmouth College
Lebanon, NH 03756

STUDY STATISTICIAN:

Hajime Uno, PhD
Dana-Farber Cancer Institute
Boston, MA 02215

FUNDING:

National Cancer Institute, UM1CA233080.

SIGNATURE PAGE

Protocol Title:	SIMPRO Research Center: Integration and Implementation of PROs for Symptom Management in Oncology Practice.
Sponsor Protocol Number:	DF/HCC IRB #18-986R/18-734
sIRB Protocol Number:	WIRB Tracking #20182593; Study #1248093
SAP Version:	Addendum to Version 2.0
Date:	June 12, 2024

Michael Hassett, MD, MPH Study Chair, Coordinating Center PI	Date
Deborah Schrag, MD, MPH Co-Chair, Co-PI	Date
Raymond Osarogiagbon, MD Co-Chair, Co-PI	Date
Sandra Wong, MD Co-Chair, Co-PI	Date
Hajime Uno, PhD Study Statistician	Date

Introduction

This addendum outlines modifications to the Statistical Analysis Plan (SAP) version 2.0 (January 9, 2024) for the pragmatic stepped-wedge cluster randomized trial at SIMPRO Research Center due to challenges encountered in the original analysis plan.

In this study, follow-up for clinical outcomes for each patient starts from the occurrence of a trigger event (i.e., initiation of a new chemotherapy for Medical Oncology and discharge after receiving a qualifying surgical procedure for Surgery). Therefore, we potentially have more than one data point (i.e., episode) from one subject in this study.

The primary analysis has been adjusted from including all episodes observed from each patient during the study period (see Pages 12 of the SAP ver. 2.0) to including only the first episode observed from each patient. Accordingly, the primary analysis has been changed from including subjects as random effects in generalized linear mixed effects models to take account of within-subject correlation (see Page 17 on the SAP ver. 2.0) to using genialized linear models because the analysis data will not involve within-subject correlation.

This change is necessitated by the observation that the majority of patients did not experience multiple episodes, leading to convergence issues with the generalized linear mixed-effects models initially planned as the primary analysis. Especially in Surgery cohort, only 2.8% subjects had multiple episodes under the intervention condition.

Changes to the Statistical Analysis Plan

1. Multiple episodes for medical record data (page 12)

Original:
- The primary analysis will include all these episodes other than the contaminated episodes.
- As sensitivity analyses, we will perform the analyses using the data consisting of only the first episode from each subject.

Revised:
- The primary analysis will include only the first episode from each subject.
- As sensitivity analyses, we will perform the analyses include all episodes other than the contaminated episodes.

2. Analysis Populations (page 15)

Original:
- 1) The primary analysis population consists of all eligible episodes under the intervention condition (A+B) (i.e., eSyM exposed) and all eligible episodes under the control condition (C) (i.e., eSyM unexposed).

- 2) The secondary analysis population consists of the episodes where patients engaged with eSyM (A) and the matched episodes during the pre-intervention period (C*).

Revised:

- 1) The primary analysis population consists of all 1st eligible episodes under the intervention condition (A+B) (i.e., eSyM exposed) and all 1st eligible episodes under the control condition (C) (i.e., eSyM unexposed).
- 2) The secondary analysis population consists of the 1st episodes where patients engaged with eSyM (A) and the matched 1st episodes during the pre-intervention period (C*).

3. Patient demographic and disease characteristics (Page 16)

Original:

- The unit of the analysis is not patient but episode. All eligible episodes eSyM- episodes will be included in this analysis.

Revised:

- The unit of the analysis is patient in the primary analysis. All eligible 1st episodes will be included in this analysis.

4. Multilevel Generalized Linear Regression Analysis (Page 16)

Original:

- The primary analysis for the outcomes in Aim 2 (Table 5) will be performed via multilevel generalized linear regression models (or generalized linear mixed-effects model; GLMM).
- Subjects
    - We will potentially have more than one data point from one subject in this study.
    - To take account of within-subject correlation, subjects will be included as random effects.
    - The primary analysis will consider random intercepts only.

Revised:

- The primary analysis for the outcomes in Aim 2 (Table 5) will be performed via generalized linear regression models (or generalized linear model; GLIM).

- Subjects
    - We will potentially have more than one episode from one subject in this study. However, only the 1st episode from each patient will be included in the primary analysis.
    - If possible, as a sensitivity analysis, we will perform multilevel generalized linear regression models (or generalized linear mixed-effects model; GLMM) by including all episodes. To take account of within-subject correlation, subjects will be included as random effects.

5. Sensitivity analyses (Page 19):

Original:

- Analysis with data from the first episode
    - In the primary analysis, all episodes in the analyses except for eSyM- episodes contaminated with the eSyM intervention will be included. We will perform sensitivity analyses using only the first episode from each patient. In these sensitivity analyses, each patient will serve only one data point in the analysis.

Revised:

- Analysis with data from all episodes
    - In the primary analysis, only the first episode from each patient will be used. In the primary analysis, each patient will serve only one data point in the analysis. We will perform sensitivity analyses using all episodes in the analyses except for eSyM- episodes contaminated with the eSyM intervention will be included.

Implementation and Documentation

These changes will be implemented immediately and documented in all relevant study documents and communications. The SAP addendum will be reviewed and approved by the IRB. All analyses will be conducted based on this SAP addendum.