Family History and Cancer Risk Study
NCT NCT05079334
IRB 201202
SAP
11/3/25

Statistical Analysis Plan

Project Title: *Outcomes of the FOREST Study: Evaluating the Impact of Patient-Directed Family History Input on Hereditary Cancer Risk Identification*

I. OVERVIEW

1. Background

Background: Hereditary cancer syndromes cause a high lifetime risk of early, aggressive cancers. Early recognition of individuals at risk can allow risk-reducing interventions that improve morbidity and mortality. Family health history applications that gather data directly from patients could alleviate barriers to risk assessment in the clinical appointment, such as lack of provider knowledge of genetics guidelines and limited time in the clinical appointment. New approaches allow linking these applications to patient health portals and their electronic health records (EHRs), offering an end-to-end solution for patient-input family history information and risk result clinical decision support for their provider.

Methods: We describe the design of the first large-scale evaluation of an EHR-integrable, patient-facing family history software platform based on the Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources (SMART on FHIR) standard. In our study, we leverage an established implementation science framework to evaluate the success of our model to facilitate scalable, systematic risk assessment for hereditary cancers in diverse clinical environments in a large pragmatic study at two sites.

Conclusions: Our research study will provide evidence regarding a new care delivery model that is scalable and sustainable for a variety of medical centers and clinics.

II. GOALS AND POPULATION

2. Analytic Goals

a. **Aim 1:** Deploy a care delivery model that will facilitate systemic risk assessment for hereditary cancers in diverse clinical environments.
b. **Research Question:** What proportion of enrolled individuals complete the MeTree questionnaire?

3. Analytical Population(s)

a. **Reach population:** defined as the participants who were invited to use MeTree (completed baseline surveys and clicked link to use MeTree)
b. **Effectiveness population:** defined as the participants who completed MeTree and had a clinical decision support risk report generated.
c. **Genetic counseling population:** The genetic counseling population includes participants who completed MeTree, received a genetic counseling recommendation and completed the genetic counseling appointment. A random selection of non-study participants with genetic counseling appointments at the same clinic and in the same time frame will be selected, 1:1.
d. See the protocol paper for additional details of these populations: https://www.sciencedirect.com/science/article/pii/S1551714424002970#f0010
e. Study participants had to meet the following inclusion and exclusion criteria to be enrolled in the FOREST study:
    i. Inclusion
        1. Site: VUMC, MMC
        2. Demographics: Age 18 years or older
        3. Able to read and communicate in English
        4. Willing and able to use the internet to report FHH information
        5. Current user of patient portal (VUMC specific)
    ii. Exclusion
        1. Not a patient at VUMC or MMC sites
        2. Diagnosed with a terminal illness
        3. Unable to speak/read English
        4. Unwilling or unable to use the Internet to report their FHH information
        5. Report previous genetic testing and/or counseling through the VUMC Hereditary Cancer Clinic
f. Analyses involving EHR -related information will be restricted to VUMC participants due to the specificity of the EHR-based outcomes.

III. OUTCOMES AND DATA COLLECTION

1. Primary outcome derivations and definitions

a. Completion of MeTree, occurs when a participant provides sufficient information to MeTree to generate a clinical decisions risk support report that is captured and logged in the study database

2. Secondary outcomes derivations and definitions

a. EHR-based indication for increased cancer risk, defined separate for the time periods
   i. EHR-based variables
      1. Variable 1: Personal history code in EHR [Yes/No]
      2. Variable 2: Family history code in EHR [Yes/No]
      3. Variable 3: Other code in EHR [Yes/No]
   ii. Derived outcomes from the EHR-based variables
      1. Variable 4: number of Variables 1-3 with value = Yes
      2. Variable 5: at least 1 of Variables 1-3 is Yes
      3. Variable 6: number of Variables 1 and 2 with value = Yes
      4. Variable 7: at least 1 of Variables 1 and 2 is Yes
   iii. Each of the variables above are defined for two time periods
      1. T0 - T1 = before consent through date of consent into the study
      2. T1 - T2 = after consent through the end of the study
   iv. Endpoint: Variable 7, at least one of personal history code or family history code, Variables 1 and 2 respectively, are met
   v. Sensitivity Endpoints:
      1. Variable 1: Personal history code in EHR
      2. Variable 2: Family history code in EHR

b. Genetic counseling appointment times
   i. Length of genetic counseling appointments

3. Handling of Missing Data

Data will be analyzed as captured in the redcap surveys and in the MeTree database. Missing data will not be imputed. In the MeTree relative conditions data, the absence of a condition will be assumed to be "no" if the health history is not set to "unknown." If health history is unknown, the condition will be set to "NA." Analysis reports will be programmatically generated from processed MeTree data cuts.

IV. STATISTICAL ANALYSIS

1. Analysis Objectives and Tasks

a. Describe the Reach populations
   i. **Analysis:** Descriptive statistics (counts and percentages, mean, median, standard deviation, IQR range) will be used to describe the demographics and baseline characteristics of the **Reach population**. Statistical tests: fisher's exact test, chi-square test, or wilcoxon ranked sum test will be used to identify differences between the participants who completed MeTree (Effectiveness population) and did not complete MeTree within the Reach population

b. Describe the genetic counseling populations
   i. **Analysis:** Descriptive statistics (mean, standard deviation, median and IQR or counts and percentages) will be used to describe differences in demographics by intervention group, MeTree and standard of care. Difference between groups will be assed via Wilcox rank sum test or fisher's exact test.

2. Primary Endpoint: Completion of MeTree

a. Implementation endpoint: Report the number and percentage of participants who completed MeTree out of the reach population. As single-arm implementation endpoint, no inferential statistics will be used.

3. Secondary Endpoint 1: EHR-based identification of hereditary risk pre- and post-MeTree

a. **Hypothesis:**
   i. **H_0:** MeTree has no impact on identification of patients at high risk for hereditary cancer
   ii. **H_A:** MeTree impacts the identification of patients at high risk for hereditary cancer

b. **Analysis:** Descriptive statistics (counts and percentages) will be used to describe the EHR-based indication for increased cancer risk (Variable 7: presence of a personal or family history codes) for the two study time periods: pre-study consent period (T0-T1) to the post-study consent period (T1-T2) within the effectiveness population. Contrast the EHR-based determination of high risk for cancer (Variable 7), between the pre-study consent period (T0-T1) to the post-study consent period (T1-T2) using the McNemar test for paired observations, test statistic and p-value will be reported.

4. **Secondary Endpoint 2:** Genetic counseling appointment length for participants who used MeTree and for randomly selected patients from the same clinic who did not use MeTree.
    a. **Hypothesis:**
        i. **H_0:** Using MeTree prior to genetic counseling has no impact on genetic counseling appointment length
        ii. **H_A:** Using MeTree prior to genetic counseling affects genetic counseling appointment lengths
    b. **Analysis:** Descriptive statistics (mean, standard deviation, median and IQR or counts and percentages) will be used to summarize appointment lengths by intervention group. Differences in genetic counseling appointment length between intervention groups will be assessed using the Wilcoxon rank sum test, p-value will be reported

5. **Exploratory Sensitivity Analysis 1 (Reach population):**
    a. **Analysis:** Use descriptive statistics (counts and percentages) to describe the EHR-based indication for increased cancer risk (Variable 7: presence of a personal or family history codes) for the two study time periods: pre-study consent period (T0-T1) to the post-study consent period (T1-T2) in the **Reach population**. Contrast the EHR-based indication for increased cancer risk (Variable 7), between the pre-study consent period (T0-T1) to the post-study consent period (T1-T2) using the McNemar test for paired observations, test statistic and p-value will be reported.

6. **Exploratory Sensitivity Analysis 2 (Personal Hx Code):**
    a. **Analysis:** EHR-based indication for increased cancer risk based on personal history (Variable 1), contrasting the pre-study consent period (T0-T1) to the post-study consent period (T1-T2), in both the **Effectiveness and Reach populations**. McNemar test for paired observations will be used to compare high risk identification in the pre-study period to the post-study period.

7. **Exploratory Sensitivity Analysis 3 (Family Hx Code):**
    a. **Analysis:** EHR-based indication for increased cancer risk based on familial history (Variable 2), contrasting the pre-study consent period (T0-T1) to the post-study consent period (T1-T2), in both the **Effectiveness and Reach populations**. McNemar test for paired observations will be used to compare high risk identification in the pre-study period to the post-study period.