PREDICT (Personalized Risk-based Follow-up of Cervical Cancer Screening in Practice), RCT

Statistical Design and Analysis

NCT#: TBA

Version date: February 22, 2024

Aim 2. Effectiveness Analyses.

Analysis Plan/ Sample Size. Our primary analysis will be based on randomization at the practice level and intention-to-treat (ITT). All eligible, enrolled patients will be part of the ITT cohort. We expect that a small number of patients may change primary care practices within our systems. These individuals will be evaluated according to their initial practice's intervention status. Prior to analysis, all data will be examined for accuracy, logical consistency, and missing data. We expect complete data for intervention status and our outcomes. We will use multiple imputation, using the MI procedure with chained equations (MICE) in SAS, to account for any missing values of covariates.[158] Through the stratified randomization, we will attempt to balance patient, practitioner, and practice characteristics between study groups (see "Intervention Design and Randomization Scheme" in section C.6 of the research strategy). However, since randomization occurs at the practice level, we will be cautious and still compare the patient, practitioner, and practice characteristics using Fisher exact tests, analysis of variance, and Kruskal-Wallis tests as appropriate to the covariate distribution. Any characteristics that show substantial clinical or statistical differences ($p<.05$) will be entered into the regression model and retained as covariates if they alter the effect estimates of the interventions by >20%.

The primary analysis model will be a random effects logistic regression, implemented through the SAS Glimmix procedure. Timely follow-up (yes/no receipt of follow-up within 120 days of the due date) will be the patient-level outcome and random effects for primary care practices and practitioners will allow for exchangeable correlation between patients within the same practice and patients seen by the same practitioner. The primary fixed predictors will be 2 indicator variables representing the 2 intervention arms and we will use a global likelihood ratio test to compare the 3 study arms (i.e., 2 intervention arms and usual care). If the global test is significant ($p<0.05$), we will compare the intervention arms to the control group and to each other (our primary comparisons with a Bonferroni-adjusted significance level of 0.0167). Comparisons to the control arm will capture the cumulative effects of the interventions, while comparisons between the intervention arms will capture the marginal effects of each level of intervention. We will include covariates for the risk-based clinical action threshold, and any patient, practitioner, or practice characteristics that are identified as confounders. Results will be presented as adjusted follow-up rates, with 95% confidence intervals, calculated using marginal standardization.

Secondary analyses will model time-to-follow-up, using a clustered proportional hazards regression to examine whether follow-up occurs as quickly as possible. Patient-level time-to-follow-up will be recorded with censoring at the end of study for patients who never received follow-up, and with death treated as a competing risk. Patients will be clustered within practitioners using the generalized estimating equation approach, implemented as a "frailty" analysis in the SAS Phreg procedure.[159] An additional correlation component for patients within practices cannot be included, but we will evaluate the robustness of our findings by using an alternative model, clustering by practice rather than practitioner. Predictors and covariates in this model will be identified in the same way as described above. The proportional hazards assumption for the intervention effects will be verified by entering a time-varying version of those predictors. Other secondary outcomes based on patient, practitioner, and team surveys will use clustered linear regression models to compare satisfaction scores between study arms. Our secondary system outcome, the number of patient contacts, will be compared between arms using clustered Poisson regression, allowing for over-dispersion. The model building will be analogous to the approach detailed above.

We will also consider several exploratory analyses. In particular, we will perform our primary and secondary analyses within each of the risk-based clinical action thresholds defined by the ASCCP guidelines. We will also calculate the Area Deprivation Index (ADI) at the block group level, based on each patient's home address and group the patients into 4 quartiles. Indicator variables for the ADI quartiles will be entered into the primary models to allow us to determine whether residence in an area with increased social vulnerability is associated with worse follow-up; and interactions between quartiles of ADI and intervention arms will indicate whether the interventions help reduce existing gaps in follow-up care. Prior to any of these subgroup analyses, we will put the appropriate interaction terms into the primary models above. However, it was not practical to design this study to have sufficient power to pursue all of these possibilities. Instead, any "findings" in these secondary analyses will be considered as clues to be pursued in future studies and in alternative databases.

Finally, while it would be clinically important to pursue cervical squamous intraepithelial neoplasia grade 3 or more severe diagnoses (CIN3+) and cervical cancer detection rates, we will not do so here because we lack

statistical power and because of the possibility that cases may be diagnosed more often in intervention than control arms because of differentially higher follow-up rates with more diagnostic evaluations. We will not have long enough follow-up times to look at differences in cancer incidence related to inadequate follow-up. Instead, we will perform descriptive analyses within each study arm to look at cases identified.

Sample Size / Power Calculations. Our primary effectiveness outcome is powered to be able to test the effectiveness of our intervention on the highest risk population, those with at least a 0.6% 5-year risk of cervical cancer who need 1 year Pap HPV or colposcopy (see Figure 4 in Research Strategy). Our primary goal is to enroll 7500 of these "higher risk" patients seen by 600 providers and randomized to 47 practices. Even after adjusting for within-practice and within-clinician correlation, this design will provide 90% power to detect an 8.4% improvement in the follow-up rate for an intervention arm compared to our control (**Table)**. Although not powered to find interactions, the Table also shows the follow-up rates that would be detectable with 80% power for different other populations at above average risk for cervical cancer (i.e., "moderate risk" abnormal results that require 3-year Pap/ HPV and women 65 years and above who have prior abnormalities that preclude screening exit. Effects of IT interventions of this size have been seen in our past studies and are reasonable to expect.[8, 77, 93, 94] Smaller effects would have minimal clinical importance.

Patient Population	**Sample Size / Effective Sample Size***	**Control Follow-up Rate****	**Intervention Follow-up Rate*****	**Power********
All higher-risk patients (90% power)	7500 / 2814	50%	> 58.4%	90%
All higher-risk patients (80% power)	7500 / 2814	50%	> 57.3%	80%
Highest risk patients needing colposcopy/ treatment	3000 / 1838	50%	> 58.0%	80%
Higher risk patients needing 1-year Follow-Up	4500 / 2300	50%	> 57.1%	80%
"Moderate risk" patients who require 3-year surveillance Testing	2000 / 1425	50%	> 59.0%	80%
Age 65+ in need of surveillance	1500 / 1155	50%	> 60.0%	80%

* Sample size was effectively reduced assuming a within-clinician correlation of 0.01 and a within-practice correlation of 0.01, as measured in the prior **mFOCUS** trial

** Assumed control rate of 50% provides the most conservative estimates of power and this approximates the observed follow-up in prior studies [5,8]

*** Follow-up rate in the intervention group would need to be at least this high to be detected with the given power

**** Power calculated assuming a Bonferroni-adjusted type 1 error of 0.0167 for the primary "all patient" analyses, and an unadjusted type 1 error of 0.05 for secondary subgroup analyses

NOTES: "Higher risk" patients are defined as those with at least a 0.6% risk of developing colposcopy in 5 years based on their prior screening results. "Moderate risk" patients are those with an above average risk of developing cervical cancer in the next 5 years (0.2% - 0.5%) or women 65+ who have prior abnormal results that require surveillance such that they cannot exit screening at age 65.

Aim 3 Implementation Analyses.

Guided by Proctor's implementation evaluation recommendations, RE-AIM QuEST and CFIR, we will assess implementation outcomes using a rigorous convergent parallel mixed methods design, in which data are collected around the same timepoints, analyzed independently, and then compared and combined.[160] This consists of qualitative and quantitative data to assess intervention reach, acceptability, fidelity/engagement, adoption, implementation and maintenance/sustainability.[161] We will use data from different qualitative sources (data triangulation) to strengthen implementation and intervention sustainability.

Qualitative Data Collection, Coding, and Analysis. We will collect qualitative data in the context of a sequential explanatory mixed methods design to explore PCP, and practice experiences around the follow-up of CCSM. We will conduct qualitative interview with mixed role groups (practitioners, administrators), balanced by network and study arm, to capture in depth impressions of the CCSM process. Qualitative exploration of

these issues allows for in-depth discussion of experience-based perceptions of care processes. Qualitative data will be used to understand/ expand upon quantitative findings and identify barriers/ facilitators to adoption, implementation and maintenance. For example, quantitative findings may indicate disparities in follow-up based on race/ethnicity or SES; themes from our qualitative data may elucidate potential biases or communication issues that are associated with these disparities. Up to four focus groups will follow a semi-structured discussion guide and will be led by an experienced moderator and will include 8 – 10 individuals. Participants will be provided $50 remuneration for participating in the groups.

Qualitative data will be recorded, transcribed, and analyzed. We will use the Rapid Qualitative Analysis Technique Assessment Process to evaluate qualitative data.[114] This methodology includes transcribing each interview, reviewing each transcript, identifying a neutral domain specific to each interview question, extracting data from each interview into a summary template form, and organizing the data from the summary template forms into a site matrix. Main topics ("domains") will be drawn from the interview and focus group guides and a summary template will be developed. Two team members will use the template to summarize the same transcripts to ensure that the domains are identifiable in the data and that there is consistency across team members in capturing the domains. Once consistency of summary content is established, transcripts will be divided up across the team and summarized using the template. Bullet points from the summary templates will then be placed into a matrix to analyze the depth and breadth of information for each domain. An audit trail of all qualitative data will be kept throughout the study.