Study Title: Promoting informed decisions about colorectal cancer screening in older adults (PRIMED Study): A cluster randomized trial

Document Title: Design and Analysis plan for the cluster randomized trial in the PRIMED Study

PIs: Karen Sepucha, PhD (ksepucha@mgh.harvard.edu) and Leigh Simmons, MD (lhsimmons@partners.org)

Statisticians: Yuchiao Chang, PhD; KD Valentine, PhD

Clinical Trials.gov: NCT03959696

Date: July 27, 2022

Purpose: The goal of the study was to advance our understanding of how to ensure that older adults are well-informed and involved in decisions about whether or not to continue CRC screening. Refer to the study protocol for an outline of the methods used in the study.

Design: The PRIMED study is a multi-site cluster randomized trial that enrolled primary care physicians and assigned them to Intervention or Comparator arms. The Intervention arm involved completion of a CME-provided 2-hour online training course in SDM communication skills and received an electronic reminder of patients with an upcoming appointment who are eligible for a CRC screening discussion. Comparator arm physicians received electronic reminders only. Each clinician will complete a baseline survey and a telephone-based SPI to capture baseline SDM skills before patient enrollment begins. Patients and physicians were surveyed after an eligible visit to assess outcomes. After enrollment is complete, all clinicians will participate in a debrief interview.

Physicians were recruited from internal medicine and family medicine practices affiliated with five hospital networks—three academic medical centers and two community hospitals– located in Massachusetts and Maine. They were eligible if they had at least 20 patients aged 75-85 due for a CRC screening or surveillance in their panel. Physicians were enrolled from May to August 2019.

Patients were enrolled from October 2019 through April 2021 (due to COVID-19, patients were not enrolled from March 13, 2020-May 24, 2020). Eligibility criteria for patient participants are listed in Table 1.

Table1. Eligibility for patient participants

Eligible	Ineligible
• Adults, age 76-85 at the time of the scheduled visit • Scheduled for non-urgent office visit with a participating physician during the study period • Due for discussion or confirmation regarding colorectal cancer screening (e.g. never been screened, <1 year to follow-up interval indicated on previous test).	• Prior diagnosis of colon or rectal cancer, inflammatory bowel disease or genetic disorder that raises CRC risk (e.g. hereditary non-polyposis CRC and familial adenomatous polyposis) • Unable to consent for themselves (e.g. moderate to severe dementia or other major cognitive limitations) • Physician excludes based on life threatening diagnosis (e.g. currently being treated for metastatic cancer) or other major event. • Unable to read or write in English or Spanish

Randomization and blinding: Eligible physician participants were grouped into strata based on self-reported gender, years in practice, prior exposure to SDM training (self-reported by physicians at enrollment), and site. Within each stratum, the study statistician used a computer random number generator to randomly assign each physician to either the intervention or control arm. Study staff assessed participating physician schedules every week to identify eligible patients to invite into the study to receive a survey to measure patient-reported outcomes. Patients who completed a survey were followed after 12 months to track colon cancer screening tests completed and a subset were eligible for a telephone interview. The PIs and the project manager were not blinded to physicians' assignment as they needed to arrange access to training

course. The biostatistician was blinded to study assignment and received a de-identified list of participating physicians to generate the assignments. Research staff were blinded both to the name and randomization status of the physician when coding the SPI. Research staff who entered the patient survey data into REDCap and the biostatistician conducting the analyses were blinded.

Patient Outcomes:

- Screening uptake: study staff examined medical records to determine receipt of any colorectal cancer screening test for all subjects within 12 months of the index visit.

The following measures were collected in the patient survey:

- SDM Process Scale: This 4-item measure assessed the discussion of 1) stopping screening as an option, 2) reasons to screen, 3) reasons not to screen, and 4) patients' screening preference. Individual items were summed to generate a total score (0-4), with higher scores indicating greater shared decision making. Score was calculated if all 4 items were answered and patients who indicated no discussion of CRC received a score of 0.
- Risk perceptions: One item assessed affective risk perception, or cancer worry. This item will be adapted from the National Cancer Institute's Health Information National Trends Survey (HINTS).
- Knowledge: Seven multiple choice knowledge items, adapted from the Colorectal Cancer Screening Decision Quality Instrument, were scored. A total knowledge score (0-100%) were calculated if at least half of the knowledge items (4 or more) were answered with the missing items considered as incorrect.

- Screening preference: One item asked patient's preferred approach to screening (with responses: colonoscopy, stool-based test, no screening, and not sure).
- Overall health: patients self-reported overall physical and mental health *(*PROMIS Scale v1.2-Global Health Physical 2a (poor to excellent).
- Patient's screening intention: One item assessed how likely the patient was to follow through with their preferred approach on a 5-point scale from Definitely will to Definitely will not.
- Screening recommendation and time spent: One item will assess the physician's recommendation about CRC screening (colonoscopy, stool-based test, no further testing, other) and one item will assess if time was spent discussing CRC screening in the visit (yes/no) and if discussed, how much time was spent (<2 minutes, 2-5 minutes, > 5 minutes).
- Single item literacy screener: reported as percentage who never or rarely require help reading medical information indicating high health literacy
- Satisfaction: One item asked "Overall, how satisfied were you with the visit" on a 4-point scale from Extremely satisfied to Not at all satisfied.

Additional data were collected to describe the sample including

- Demographics: race, ethnicity, employment, marital status and education
- Family history of colorectal cancer and personal history of prior polyp removal.
- Patients' attitude: self-report of importance of finding cancer early, their feelings regarding the risk & benefits of screening for colon cancer, how much they believe screening will help, single-item maximizer/minimizer scale, tolerance of uncertainty

Physician Outcomes:

Physicians completed short background and baseline surveys to collect demographics and prior SDM training experience. They were also asked to review a list of their patients before being randomized into the study. Further, physicians completed a telephone-based SPI before staff started to enroll their patients. Lastly, once patient enrollment was complete, all physicians participated in a debrief interview.

- Background physician survey: Items assessed the number of years the physician had been in clinical practice, years working at their current site, and clinical sessions each week, as well as age, race, ethnicity, specialty, gender, and contact information. Additionally, physicians reported any prior formal training in SDM (and defined what that experience was).
- Baseline physician survey: Items assessed attitudes towards colonoscopy, stool-based tests, and not screening patients in this age group, subjective norms, and behavioral intention to engage their older adults in decisions about CRC screening. Physicians also completed the knowledge items and the physician's reactions to uncertainty scale.
- Simulated patient interaction: Physicians in both arms completed a telephone-based SPI before patient enrollment started. Intervention arm physicians complete the SPI after the SDM training. The SPIs were audio recorded and professionally transcribed, de-identified, and coded by two trained coders, who were blinded to the physicians' assignment. Coding followed the well-validated Braddock's Informed Decision Making framework to assess for presence of SDM elements.
- Post-visit survey: Physician participants completed a short online survey via REDCap after each eligible patient visit. The survey contained the SDM Process items adapted for physicians as well as an opportunity to provide a reason for not discussing CRC screening.

- Debrief interview: After patient recruitment was complete, research staff conducted a brief interview with participating physicians that followed a structured interview guide to assess physicians' attitudes toward SDM, their perceptions of the study, and satisfaction with the intervention, and ideas for improvement.

Sample Size: The study was powered to detect a small to medium effect size difference in the primary outcome, SDM process score. With 500 surveys (250 per arm), assuming an intraclass correlation coefficient of 0.03, the effective patient sample size was estimated at 394, which would enable detection of a difference of 0.28 standard deviations with 80% power and a two-sided significance level of 0.05.

Statistical methods: Responders and non-responders were compared to examine potential non-response bias using two-sample t-tests or chi-square tests. Patient sample characteristics were compared between arms using two-sample t-tests or chi-square tests. Multivariable regression models were used to adjust for potential effects of any unbalanced variables.

The following hypotheses were evaluated using an intention to treat approach, and patient outcomes were analyzed based on their physicians' assigned arm regardless of whether the physician completed the training, received the reminder, or discussed CRC screening.

1. Examine the effects of the interventions on **patients' SDM Process scores and knowledge score**:

a. Compared to the Comparator group, patients seen by physicians in the Intervention arm would report higher SDM Process scores (primary outcome) : We will examine the distribution of the SDM Process score and apply variable transformation techniques if necessary. We will use linear regression models with the Generalized Estimating Equations (GEE) techniques to account for the patients within clinician data structure. We will be able to difference of 0.28 SD for the SDM Process score with 80% power. Studies using the SDM Process survey have found effect sizes ranging from 0.39SD – 0.88SD when comparing sites that used have formal decision support (coaching or decision aids) and those that did not. Although we have sufficient power to detect a smaller difference, we are interested in a meaningful difference 0.4SD or higher.

b. Compared to the Comparator group, patients seen by physicians in the Intervention arm would have greater knowledge: We will use linear regression with GEE to compare the mean knowledge score between the two arms. The target sample size will provide 80% power to detect a difference of 0.28 SD for the knowledge scores after adjusting for the effects of clustering. Although there are no standards for a minimally important difference in knowledge, SDM studies have found effect sizes of 0.25SD-0.8SD in studies of decision aids, and this study is adequately powered to detect a meaningful difference in that range, and

2. Examine the effects of the interventions **on patients' preferences for screening**, the extent to which patients receive their preferred approach to screening, and on **CRC screening rates**.

a. Compared to the Comparator group, patients seen by physicians in the Intervention arm will receive their preferred approach to screening (either colonoscopy, stool-based test, or no further screening). For these analyses, we will compare the percentages of patients who prefer to stop screening, the percentages of informed patients who received their preferred screening in the 12 months after the visit, and overall rates of screening across the two groups using logistic regression model with the GEE approach to adjust for clustering of patients within clinicians. We will have 81% power to detect a difference of 14% in rates (e.g. decrease from 61% to 47%).

b. Both interventions will reduce screening rates compared to concurrent controls (rates of clinicians not involved in study). The screening data will be available for all patients from all participating sites. We estimate a total of 6248 patients will be eligible for screening from 436 clinicians: 5353 patients from clinicians not participating the study and 448 from each arm of clinicians participating the study. After taking into account clustering effect, we will have at least 86% power to detect a 6% decrease in CRC screening rates of patients from clinicians participating in the study compared to patients from clinicians not participating.

3. Examine the effects of the interventions on **physicians' confidence and demonstration in their SDM skills** and their **satisfaction**.

a. Compared to the Comparator group, physicians in the Intervention arm will report higher confidence in their SDM skills: Clinician confidence will first be analyzed as a continuous variable. Variable transformation will be performed to improve normality assumption if deemed necessary. Additionally, we will dichotomize the

variable into very or extremely confident vs. not. A two-sample t-test or Wilcoxon rank sum test, as appropriate, will be used to compare the continuous outcome while a chi square test will be used to compare the dichotomized outcome. With a sample size of 50 clinicians, we will have 81% power to detect difference of 35%, e.g. from 50% to 85%, in the percentage of clinicians who are very or extremely confident in different elements needed to conduct SDM conversations. In our prior study of the SDM skills webinar, we found a 30-40% absolute increase in clinician confidence.

b. Compared to the Comparator group, physicians in the Intervention arm will demonstrate more SDM skills in the SPI: We will code each transcript using Braddock's Informed Decision Making framework that covers core aspects of SDM (scores range from 0 to 9). First, we will determine whether a two-sample t-test or a Wilcoxon rank sum test is more appropriate to compare the two groups. Assuming the score is normally distributed, a sample size of 25 in each group will have 80% power to detect a 0.81SD difference in the mean Braddock score (or about 1.4 points out of 9). Our prior study of a 3-hour CME online SDM skills webinar, we found an increase of 1.2SD on the Braddock scale after training.

c. Compared to the Comparator group, physicians in the Intervention arm will report higher satisfaction with the visit. The percentage of clinicians who report that they are extremely satisfied with the visit will be compared across arms using a chi-square test.

4. Examine the effects of the interventions on **caregivers' SDM Process scores**: Caregivers who attended a visit with an eligible patient (and for whom the patient provided contact

information) were invited to complete a post-visit survey. They were asked to complete an adapted version of the SDM Process survey to provide their perspective on the conversation and involvement of the patient. A total score will range from 0-4, with higher scores indicating more shared decision making. These analyses are exploratory.

Heterogeneity of Treatment Effects (HTE)

The goal of the HTE analysis is to identify differential treatment effects among subgroups of patients including the following pre-specified factors:

- Physician factors included (1) hospital network, (2) gender, (3) age, (4) years in practice and (5) prior experience with SDM training.
- Patient factors included (1) sex (2) age (3) prior screening history, and (4) overall health.

We will use linear or logistic regression models with the GEE approach to test interactions between study arms and these factors. With the unexpected disruption due to COVID, we added time period (enrollment pre and post COVID) to the HTE analyses post hoc. Due to the exploratory nature of the HTE analysis, we reported treatment effects in each subpopulation when the significance level for the interactions between intervention and these factors was ≤ 0.1.