

12-9-2015

Analytic Challenges Arising from the STOP CRC Trial: Pragmatic Solutions for Pragmatic Problems

William M. Vollmer

Kaiser Permanente, Center for Health Research, william.vollmer@kpchr.org

Beverly B. Green

Group Health Research Institute, green.b@ghc.org

Gloria D. Coronado

Kaiser Permanente, Center for Health Research, Gloria.D.Coronado@kpchr.org

Follow this and additional works at: <http://repository.edm-forum.org/egems>



Part of the [Health Services Research Commons](#)

Recommended Citation

Vollmer, William M.; Green, Beverly B.; and Coronado, Gloria D. (2015) "Analytic Challenges Arising from the STOP CRC Trial: Pragmatic Solutions for Pragmatic Problems," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 3: Iss. 1, Article 19.

DOI: <http://dx.doi.org/10.13063/2327-9214.1200>

Available at: <http://repository.edm-forum.org/egems/vol3/iss1/19>

This Methods Case Study is brought to you for free and open access by the the Publish at EDM Forum Community. It has been peer-reviewed and accepted for publication in eGEMs (Generating Evidence & Methods to improve patient outcomes).

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

Analytic Challenges Arising from the STOP CRC Trial: Pragmatic Solutions for Pragmatic Problems

Abstract

Context: Pragmatic trials lack the relatively tight quality control of traditional efficacy studies and hence may pose added analytic challenges owing to the practical realities faced in carrying them out.

Case Description: STOP CRC is a cluster randomized trial testing the effectiveness of automated, electronic medical record (EMR)-driven strategies to raise colorectal cancer (CRC) screening rates in safety net clinics. Screen-eligible participants were accrued during year 1 and followed for 12 months (measurement window) to assess completion of a fecal screening test. Control clinics implemented the intervention in year 2.

Implementation Challenges/Analytic Issues: Due to limitations on how we could build the intervention tools, the overlap of the year 1 measurement windows with year 2 intervention rollout posed a potential for contamination of the primary outcome for control participants. In addition, a variety of factors led to a lack of synchronization of the measurement windows with actual intervention delivery. In both cases, the net impact of these factors would be to diminish the estimated impact of the intervention.

Proposed Solutions: We dealt with the overlap issue by delaying the start of intervention rollout to control clinics in year 2 by 6 months and by truncating the measurement windows for intervention and control participants at this point. In addition we formulated three sensitivity analyses to help address the issue of asynchronization.

Conclusion: This case study might help other investigators facing similar challenges think about such issues and the pros and cons of various strategies for dealing with them.

Acknowledgements

Research reported in this publication was supported by the National Institutes of Health through the National Center For Complementary & Alternative Medicine under Award Number UH2AT007782 and the National Cancer Institute under Award Number 4UH3CA188640-02.

Keywords

pragmatic trials, data analysis, methodology, colorectal cancer screening, cluster-randomized study, 2014 Group Health Seattle Symposium, Electronic Health Records, Quality Improvement

Disciplines

Health Services Research

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).



Analytic Challenges Arising from the STOP CRC Trial: Pragmatic Solutions for Pragmatic Problems

William M. Vollmer, PhD;^j Beverly B. Green, MD, MPH;ⁱⁱ Gloria D. Coronado, PhDⁱ

ABSTRACT

Context: Pragmatic trials lack the relatively tight quality control of traditional efficacy studies and hence may pose added analytic challenges owing to the practical realities faced in carrying them out.

Case Description: STOP CRC is a cluster randomized trial testing the effectiveness of automated, electronic medical record (EMR)-driven strategies to raise colorectal cancer (CRC) screening rates in safety net clinics. Screen-eligible participants were accrued during year 1 and followed for 12 months (measurement window) to assess completion of a fecal screening test. Control clinics implemented the intervention in year 2.

Implementation Challenges/Analytic Issues: Due to limitations on how we could build the intervention tools, the overlap of the year 1 measurement windows with year 2 intervention rollout posed a potential for contamination of the primary outcome for control participants. In addition, a variety of factors led to a lack of synchronization of the measurement windows with actual intervention delivery. In both cases, the net impact of these factors would be to diminish the estimated impact of the intervention.

Proposed Solutions: We dealt with the overlap issue by delaying the start of intervention rollout to control clinics in year 2 by 6 months and by truncating the measurement windows for intervention and control participants at this point. In addition we formulated three sensitivity analyses to help address the issue of asynchronization.

Conclusion: This case study might help other investigators facing similar challenges think about such issues and the pros and cons of various strategies for dealing with them.

^jKaiser Permanente, Center for Health Research, ⁱⁱGroup Health Research Institute

Context

Pragmatic trials (also called “effectiveness studies”) are designed to evaluate the impact of interventions when delivered in real-world settings.^{1,2} In contrast to more traditional explanatory (or efficacy) studies, which seek to evaluate the impact of an intervention under idealized circumstances, pragmatic trials are generally characterized by the following: large, representative populations selected with minimal exclusion criteria; interventions that are delivered by regular health care providers (as opposed to research personnel) in real-world clinical settings; and a lack of the relatively tight quality control that is common in efficacy studies.

Pragmatic trials can thus be difficult to carry out and often require more complex designs and statistical methods.³⁻⁵ For example, many such trials test interventions that are delivered at the level of health clinics, rather than of individual patients. This gives rise to cluster randomized designs in which whole clinics commonly serve as the unit of randomization.⁶ Included in this category of trials are stepped wedge designs, in which all clinics may receive a given intervention, but the intervention rollout is staggered over time with the order of rollout randomly determined.⁷

Apart from the added complexity of the analytic models and accompanying sample size calculations required for these types of designs,^{6,7} pragmatic trials by their very nature may pose additional analytic challenges owing to the practical realities that investigators are likely to face in carrying them out. This paper discusses several analytic challenges posed by the Strategies and Opportunities to STOP Colorectal Cancer in Priority Populations (STOP CRC) Trial and describes how we addressed them.

Case Description

Study Design

STOP CRC is a cluster randomized trial designed to test the effectiveness of automated, electronic medical record (EMR)-driven strategies to raise colorectal cancer (CRC) screening rates in safety net clinics.⁸ A total of 26 clinics were randomly assigned to receive either usual care or active intervention during the first year of intervention (study year 1), with the 13 usual care clinics adopting the intervention in study year 2 (see Figure 1). The primary analysis compares the screening rates between screen-eligible patients in the intervention and control clinics during the first year of intervention activity.

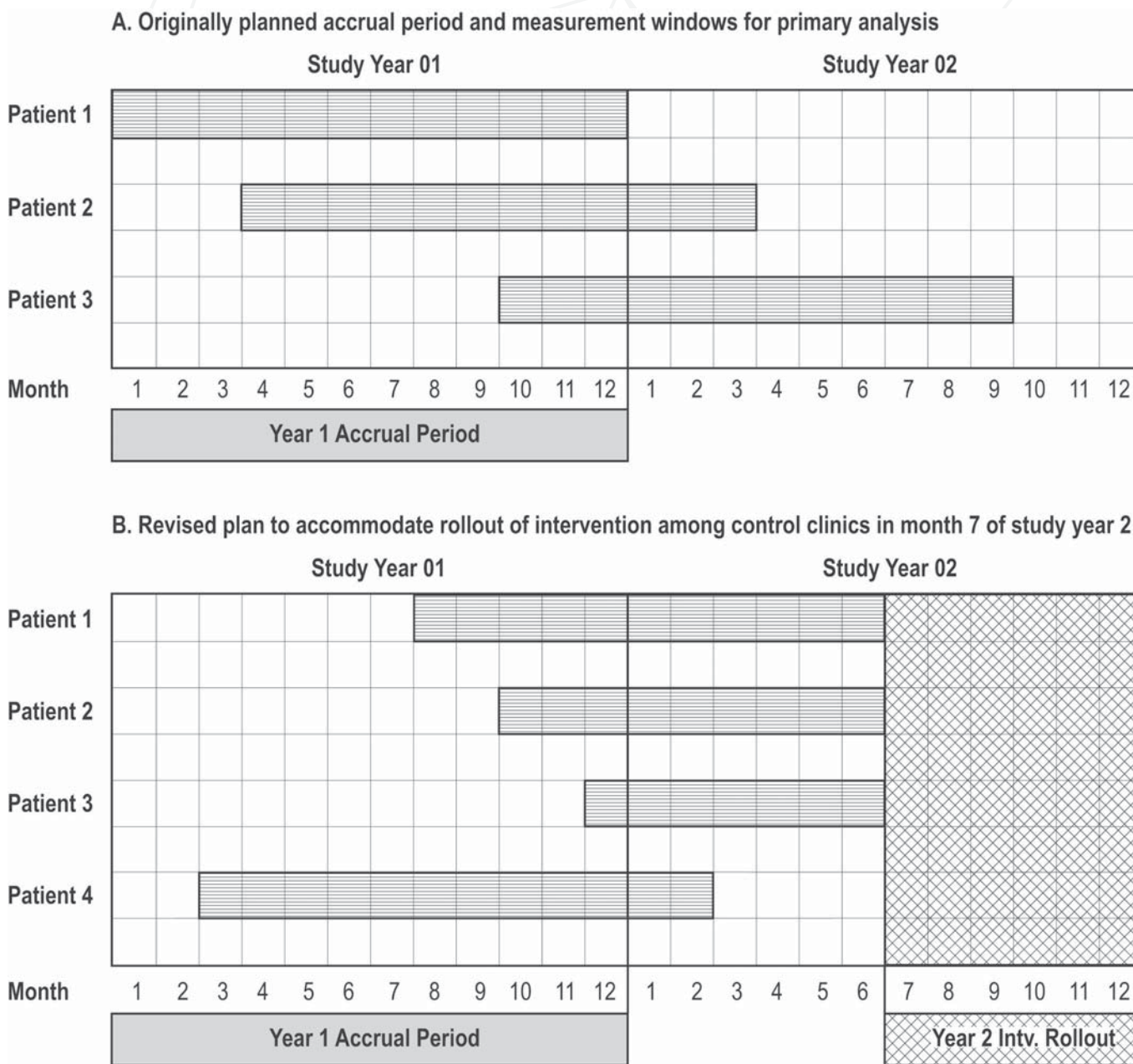
Intervention

The participating clinics share a common EMR system (EPIC) that is maintained by a central administrative site. As part of the study, these administrative staff created an EMR-embedded registry to identify individuals who are due for CRC screening based on the United States Preventive Services Task Force (USPSTF) recommendations⁹ and who did not have a gastroenterology or colonoscopy referral in the past year or an ordered fecal occult blood test (FOBT) in the prior 6 months. This information is made available to intervention clinic staff as a queryable report—or Clinical Decision Support (CDS) tool—within the EMR that can also be used to generate intervention mailings. The registry is updated nightly.

Since the participating health centers do not serve defined populations, they operationally define “active” patients as those with at least one clinic visit in the past 12 months. This rule was chosen because the National Quality Forum (NQF) metric that



Figure 1. Illustration of separate accrual and individual measurement windows



Note: Participants are accrued into the analysis sample during a common 12-month accrual period base, while return of FIT kits are assessed over individual measurement windows measured from the date of initial entry screen eligibility. Panel A depicts the original analysis plan, in which each measurement window lasted for 12 months. Panel B depicts the revised plan to truncate measurement windows at the shorter of either 12 months or the start of intervention rollout for control clinics in month 7 of study year 2.

community health centers are required to report for CRC screening rates (NQF0034) defines a “patient” as a “person with a visit in the prior 12 months.”¹⁰ We built this into the eligibility criteria for the registry and, as a result, individuals may enter and leave the registry on a dynamic basis over time.

Intervention clinic staff use the CDS tool to determine who needs to receive the intervention—which consists of a mailed introductory letter, followed by a mailed fecal immunochemical test (FIT) kit, and (if needed) a mailed reminder letter. To maximize implementation, we let the clinics determine the exact process by which they accomplish this task. This was important since we knew that there would be a large group of individuals who entered the registry at the outset of the study. Although clinic staff could choose to mail to participants as soon as they appeared on the registry, most clinics wanted to spread the work evenly over the year so as to require a more uniform staffing configuration over time. Some clinics, for example, chose to mail to participants using the “birthday” model—that is, mailing to participants only during their birth month. While the study design would have been cleaner and analytic challenges minimized had we imposed a common roll-out process across all clinics, allowing them to fit it to their own workflow was more consistent with community participatory principles, which state that if a clinic is implementing an intervention then it should control the human resources and workflow needed for implementation.¹¹ It also increases the likelihood of downstream maintenance should the intervention prove effective.

After randomization, intervention clinic staff were trained to use the CDS tool to deliver the intervention. They were also trained in best practices to update their medical records with prior CRC screening events. Clinics varied in whether and how much chart cleaning they performed.

Participant Accrual and Follow-Up

Individuals are accrued into the primary analysis sample for one year from the date of randomization, with the date of accrual being the first day on which each individual met criteria for the previously described registry. We refer to this one-year period as the “accrual window.” For convenience we shall also refer to this first year postrandomization as “study year 1,” and the subsequent year as “study year 2.” Identical rules define the analysis samples for both intervention and control-clinic participants.

Once accrued, participants are followed over time from the date of accrual to see if they complete a FIT. The original analysis plan proposed to follow individuals for 12 months from their date of accrual to determine if a FIT was completed. For each participant we refer to this period as that individual’s “measurement window.” Other than for participants who were initially accrued into the study on the first day of randomization, the accrual and measurement windows do not overlap (Figure 1, panel A). For those individuals accrued on the last day of study year 1, their measurement windows extend through the entirety of study year 2.

Although we know from previous studies that most patients will return their FIT kit within three months of receipt,¹² we adopted the 12-month measurement window in part because we knew that not all patients would receive their kits right after entry into the registry (e.g., the birthday model described above). The implicit assumption was that over the course of a year all eligible individuals would be sent a FIT kit.

Primary Analysis

Because the primary analysis focuses on clinics, not individual patients, as the unit of analysis, the primary outcome variable is the proportion of screen-eligible patients at each clinic who complete a FIT. Treating these observations as approximately



normally distributed, we will use a mixed model analysis of covariance (ANCOVA) to estimate the screening probabilities as a function of intervention while adjusting for baseline clinic screening rate.

Implementation Challenges and Resulting Analytic Issues

During study implementation we encountered a variety of challenges that threatened the validity of our primary analysis. While the nature of the challenges we encountered are not unique to pragmatic trials, we believe they are likely to be more common in such trials due to both the types of designs commonly used in such studies and the challenges of implementing system-based interventions within freestanding health clinics.

Overlap of Measurement Window with Year 2 Intervention Rollout

For control clinic participants, the overlap of the year 1 measurement windows with study year 2 posed a potential for contamination since we initially planned to roll out the intervention for the control clinics at the beginning of study year 2. Although the overlap problem could easily have been avoided by including control clinic patients on the CDS tool only if they were past their year 1 measurement window, practical constraints encountered during implementation precluded this. Even if it were possible to develop the needed filter, the clinics were reluctant to invest limited staff resources to develop a separate tool for control clinics that would not be used long-term and would not let control clinics see complete lists of screen-eligible patients at any given point. These issues all resulted, at least in part, from the pragmatic nature of the trial, which dictated that the CDS tool be developed and maintained by centralized administrative staff as a clinical tool that was fully integrated into the EMR. The choice to roll out the intervention in year 2 for the control clinics was a design issue, but it was the unanticipated

challenges that surfaced while working out the implementation details that ultimately created the problem for us.

As a result of this issue, we found ourselves faced with the prospect of study year 2 intervention activity overlapping with our study year 1 measurement windows for control participants. This in turn would create an upward bias in our year 1 usual care response rates and therefore would cause us to underestimate the true treatment effect.

Delayed Intervention Implementation and other Timing Issues in Study Year 1

We originally planned to stagger randomization over time since we expected the intervention clinics to be at different stages of readiness to implement. However the clinics were very motivated to begin screening right away, in part because the state of Oregon was offering financial incentives to clinics that met certain 2014 CRC screening benchmarks for Medicaid patients. As a result, the clinics were unwilling to agree to a phased rollout process, and wanted time to plan their strategy and assign staff to carry out the intervention. Thus, we made the decision to randomize and unblind all of the clinics at once.

This decision ultimately led to sizeable lags between the randomization date and the actual intervention rollout (Table 1). In part this was due to the time needed to conduct all of the trainings, which happened sequentially clinic by clinic, but was also due to a variety of other factors that reflected the pragmatic nature of the intervention:

- Some clinics were in the process of converting from the use of FOBT kits to the FIT kits and delayed rollout until this conversion was completed.
- Final testing of the CDS tool was delayed by an upgrade of the EPIC system that rolled out about 4 months postrandomization and, hence, delayed training on the tool.

Table 1. Timeline of Intervention Rollout

	CLINIC NETWORK ¹							
	1	2	3	4	5	6	7	8
Randomization (R)	02-02-14	02-02-14	02-02-14	02-02-14	02-02-14	02-02-14	02-02-14	02-02-14
Clinic training	02-20-14	05-15-14	05-16-14	06-06-14	06-10-14	06-17-14	08-06-14	10-06-14
Testing complete²	07-15-14	06-15-14	06-15-14	10-01-14	09-27-14	08-15-14	11-15-14	01-15-15
Date of first mailing	6-27-14	6-11-14	6-16-14	9-23-14	9-29-14	7-26-14	7-9-14	1-31-15
Days from R to 1st mailing	143	127	132	231	237	172	155	361

Notes: ¹Some networks consist of more than one participating intervention clinic. ²All clinics were required to undergo a series of test protocols to confirm they were ready to have the CDS tool officially activated. For networks with multiple clinics, this was a common date for all clinics.

- Some clinics chose to invest considerable staff effort in updating their records prior to rollout so that they would accurately capture past colonoscopies and sigmoidoscopies.
- Some clinics experienced significant staff or leadership turnover, which introduced challenges in scheduling trainings.
- The influx of new patients resulting from the rollout of the Affordable Care Act meant that some clinics were overburdened and understaffed. Among other things, this delayed staff training and, on occasion, the timing of program rollout.
- Some clinic staff had no prior experience with the reporting tools we used and, hence, required extra time in training.

In the end, none of the 13 intervention clinics began intervention mailings until at least 4 months postrandomization, and one of them did not begin mailings until the very end of study year 1. In response to these delays, many clinics modified their initially planned workflows to include large

catch-up mailings near the end of 2014 to help deal with their backlogs.

The net impact of these delays was that the start of the 12-month measurement windows for most of our intervention participants was substantially out of sync with the start of the intervention mailing. This issue, if unaddressed, would cause us to underestimate the true impact of the intervention mailings once they started. Unfortunately the magnitude of these problems varied from clinic to clinic, and we have no simple way to mimic these lags for the control clinics since the process by which the intervention was rolled out was allowed to vary from clinic to clinic.

Use of a Real-Time Intervention Tool That Didn't Coincide with Static Analysis Sample

As noted previously, individual patients might enter and drop out of the registry on a dynamic basis over time. One common way this could happen was for



someone who initially met the active patient (i.e., visit within the past 12 months) criterion to no longer meet this criterion later. Thus a patient who was initially accrued into the registry on the basis of a visit 11 months prior would drop off the EMR report at the end of the 12th month if that patient didn't have another clinic visit in the meantime.

This made perfect sense from the clinic's perspective and was, we believe, the correct pragmatic choice for how to build the CDS tool. However since the primary analysis must of necessity include all individuals who ever qualified for screening, even if they did so for only a few days, it would include many intervention participants who never received (or even could have received) an intervention mailing by the time their clinic launched the intervention. Although this would have been an issue for some patients even if clinics began intervention mailings right away (depending on the rollout process the clinic chose), it was compounded by the intervention delays.

The impact this issue would have on our analysis would be to cause us to underestimate the true impact of the intervention.

Impact of the Affordable Care Act (ACA) Rollout

Due to the concurrent rollout of the ACA, many of the clinics experienced substantial increases in enrollment. In addition to the impacts noted above, this had one other important impact. In an effort to schedule initial visits for new patients as soon as possible, some clinics made the decision to defer appointments for many of their existing patients. Particularly when coupled with the implementation delays described above, these delays in scheduling routine follow-up visits for existing members further exacerbated the problem of patients appearing to no longer be active clinic members once intervention mailings were begun.

Proposed Solutions

The analytic issues raised above fall roughly into two main categories: contamination of the year 1 measurement window by year 2 intervention rollout in the control clinics, and lack of synchronization of measurement windows with actual intervention delivery. In this section we describe our planned approach to dealing with these issues. For the former problem we adopted a design-based solution, while for the latter we propose a series of sensitivity analyses. In each case a strict requirement of any possible solution was that it could be applied equally to intervention and control clinic data.

Contamination of Measurement Windows by Year 2 Intervention Rollout

In order to avoid contaminating the year 1 outcome data for control participants we adopted the following design-based solution. First, control clinics agreed to delay turning on the CDS tool until month 7 of study year 2. This effectively eliminated the issue of overlap for participants accrued through the first 6 months of study year 1 since their measurement windows would extend no further than through the end of month 6 of study year 2. Second, for participants accrued in the latter half of study year 1, we decided to truncate their measurement window at the end of month 6 of study year 2 (Figure 1, panel B). Thus for these individuals their measurement windows varied in length from 6 to 12 months, depending on their date of accrual into the analysis sample. Finally, we applied the same rules to both intervention and control clinic participants to avoid introducing any biases in how the outcomes were measured between these groups.

An alternative solution would have maintained a fixed measurement window and redefined the length of the accrual and measurement windows to be no more than 18 months. For instance we might have accrued patients for only 6 months and retained the

full 12-month measurement window, or we might have gone with a 12-month accrual window and a 6-month measurement window, or any combination between these extremes. Had we gone this route we would have opted for the 6-month accrual window and 12-month measurement window. Power would be minimally impacted by the loss of subjects (about 25 percent of the full sample) because of the cluster randomized design, while the 12-month measurement window would maximize our chance of measuring the true intervention impact despite the delayed rollout.

We opted for our proposed solution for two main reasons. First, it doesn't exclude anyone from the analysis sample. Second, the shorter measurement windows for those participants accrued later in study year 1 are likely to still be sufficient to allow us to see an intervention impact since, by that time, the lags between accrual and start-up were much smaller. We also know from other work^{12,13} that most individuals who were mailed a FIT kit will return it within 3 months. Thus even a 6-month measurement window for someone accrued at the end of study year 1 is likely to be sufficient to allow us to see an intervention impact. On the downside, the varying lengths of the measurement windows under our proposed solution do raise some ambiguity as to the meaning of our outcome probability (i.e., it is no longer the probability of observing a returned FIT kit within a fixed number of months). However given the implementation delays we experienced, the meaning of our original 12-month probabilities would have been vague anyway since they wouldn't really reflect the impact of 12 months of actual intervention rollout.

While survival analytic techniques would have enabled us to deal with the issue of right censoring of the measurement windows for our control participants, we do not believe it would have been an appropriate analytic choice for this study given our focus on clinic level data as our primary outcome of interest. Furthermore, the use of such models still

wouldn't have allowed us to deal with the issue of delayed start-up. While one might propose to use the date of intervention mailing to individual participants as time 0 for such an analysis, there is no well-defined, comparable lag that we could use for control clinic participants. We do propose to use survival analytic techniques to estimate time to return of a FIT among those individuals who are mailed a kit.

Sensitivity Analysis #1: Delay Accrual of Patients for Six Months

As an alternative to our primary analysis, we considered three sensitivity analyses that we thought might more accurately estimate the true impact of the intervention. The first of these would ignore the initial 6 months of study year 1 and effectively begin accrual of study participants 6 months later. In this case participant accrual would continue for just 6 months, concluding—as with the original analysis plan—at the end of study year 1. As before, measurement windows would be truncated after the first 6 months of study year 2 to avoid overlap with the year 2 rollout for control clinics, hence the measurement windows would again range from 6 to 12 months in duration. This approach has the advantage of greatly minimizing the lag between the start of an individual's measurement window and the actual start of intervention activity at the individual's clinic.

On the downside, this analysis requires that we totally redefine the analysis sample rather than limiting the sample to those who became screen eligible 6 months or more postrandomization. This is important since many of those who were initially accrued as screen eligible during the first 6 months of study year 1 would still have been eligible 6 months later. Due to the shortened accrual window, the number of patients accrued into the analysis sample will be less than those accrued for the primary analysis. However, since power for this



cluster randomized trial is driven largely by the number of clinics rather than by the sample size per clinic, the impact on study power should be small.

We recognize that this analysis may introduce some biases of its own since some intervention activity did happen during those initial 6 months, and that could have affected subsequent screening eligibility under the new rules. Still, we feel that it may give a more accurate estimate of the intervention's true impact.

Sensitivity Analysis #2: Redefine Visit Window for Accrual into Analysis Sample

A second sensitivity analysis would only accrue patients into the analysis sample if they had not only met USPSTF guidelines for screening, but also had a *recent* clinic visit (a period much shorter than the 12 months used to populate the registry). Otherwise it would look identical to the primary outcome analysis. Shortening the visit criterion greatly minimizes the problem of patients who are no longer deemed to be active patients when the intervention ultimately rolls out. For instance, with a 1-month visit requirement a patient would still be deemed an active clinic patient 11 months later, even in the absence of an intervening clinic visit.

The main limitations of this approach are the following: (1) it would once again require us to redefine the analysis sample; and (2) it may introduce bias of its own since some patients who would not qualify for intervention under the new rules may still have received some intervention, and this could influence subsequent eligibility.

Ultimately we will probably opt to not include this analysis—due to limited staff resources. Also, we feel that it is not likely to buy us much in comparison with the preceding analysis, which will also serve to minimize to a large extent the issue of patients who have become inactive once the intervention rolls out. Nonetheless we include it as one of the viable options we considered.

Sensitivity Analysis #3: Analysis Based on Stepped Wedge Design

Our analytic challenges result from the fact that we are trying to estimate the steady state impact of our intervention during what is essentially a start-up year, and also from the fact that our measurement and accrual windows are out of sync with one another. A third sensitivity analysis would attempt to tackle these issues head on by viewing our design as a three-period, stepped wedge design in which we seek to estimate separate start-up and steady state effects relative to usual care. We would use as our outcome measure a Healthcare Effectiveness Data Information Set¹⁴ (HEDIS)-like score, the measurement window for which coincides with the study year. Under this framework, the year prior to randomization serves as the baseline year, during which all clinics receive usual care; in our current study year 1, the intervention clinics experience the intervention start-up effect while the control clinics continue to receive usual care; and in our current study year 2, the intervention clinics experience the steady state intervention effect while the control clinics experience the start-up effect (Table 2). Since this design allows us to directly estimate start-up versus usual care effects and steady

Table 2. Conceptual Layout for Study as a Stepped Wedge Design

	BASELINE	YEAR 1	YEAR 2
Control Clinics	usual care	usual care	start-up
Intervention Clinics	usual care	start-up	steady state

state versus start-up effects, it also permits estimation (albeit with greater variance) of steady state versus usual care effects.

One question for this analysis is whether we should define the final year as beginning midway through current study year 2 (when the CDS tool will be turned on for the control clinics) or at the beginning of current study year 2. For the meaning of “start-up” to be the same in each year, we believe the final year for this analysis should be chosen to coincide with our current study year 2. A major reason for the delayed start-up in study year 1 is the training that needs to occur at each intervention site, both in the use of the new CDS tool and in strategies to update existing records with prior colonoscopies and sigmoidoscopies. This work will happen in the control clinics as well, but during the first half of study year 2. The only thing that is delayed is the turning on of the CDS tool itself. However, since most intervention sites didn’t begin using this tool in year 1 for at least 6 months anyway, in using current study year 2 for this analysis we are effectively just formalizing what happened in study year 1. Thus we believe that the final year of the stepped wedge design will provide the most valid analysis if it coincides with current study year 2.

The stepped wedge design further differs from the primary analysis and the other sensitivity analysis in that the analysis is not limited to individuals who are not current in their CRC screening. Rather it includes all age-eligible members for whom screening is recommended by the USPSTF, including individuals currently covered by prior colonoscopies or sigmoidoscopies who are not eligible for our intervention mailings. As such, it should tend to dilute the impact of our intervention to some extent. The HEDIS-like score used for this analysis also uses somewhat different visit eligibility requirements than will be used for our other analyses. Nonetheless, it is a policy-relevant metric and, hence, has strong appeal.

Conclusion

Issues arising during the implementation of the STOP CRC pragmatic clinical trial have threatened the validity of our a priori, specified primary-analysis strategy and forced us to modify our design somewhat and to consider a number of sensitivity analyses not originally planned that we feel might provide a more accurate estimate of the impact of our intervention. While several of the issues we have discussed may be unique to our study, we believe that they may be more broadly representative of the types of issues that may arise in the conduct of large pragmatic trials, particularly those in which the interventions are embedded in health care delivery systems. While such issues may occur in efficacy studies, we believe they are likely to occur more frequently in pragmatic trials owing to the types of designs commonly used and the looser control one typically finds in pragmatic trials.

It is possible that at least some of the challenges we faced could have been foreseen and avoided by an alternative design or analysis strategy. Perhaps the most obvious example is the contamination of the year 1 measurement windows for control participants by the rollout of the intervention in these clinics in year 2. This situation is not uncommon for pragmatic trials of health care system interventions, where the use of staggered implementation, including formal stepped wedge designs, are often employed. We also believed that almost all eligible patients would show up on day one, and we did not anticipate the timing or impact new enrollees would have on our design. Also, although the project statistician was aware of the potential problem, he believed that a simple solution lay in filtering control participants from the CDS tool until after their year 1 measurement window had lapsed. Unfortunately he was not involved in operational discussions about how the CDS tool was constructed and didn’t appreciate the difficulty we would face in trying to



implement this filter. By the time he raised the issue, the tool had already been built and it was too late to change it, forcing the delayed rollout in year 2 as at least a partial solution. In retrospect the problem might have been avoided had the project statistician been more actively involved in operational issues about how the intervention tools were being built or had he raised the issue earlier with other study staff.

The delays in intervention implementation were anticipated to some extent, although they wound up being longer than we had expected (owing in part to the concurrent rollout of both the ACA and a new EPIC upgrade, and the need for some clinics to convert from FOBT to FIT kits). We had originally hoped to stratify randomization based in part on each clinic's anticipated readiness to implement. However several factors, related at least in part to the pragmatic nature of the trial, intervened. First and foremost, this was a cluster randomized trial with a relatively small number of clinics that differed on a number of factors likely to be related to our outcome. Traditional randomization stratified on a number of factors was, thus, not feasible. We considered constrained randomization but eventually opted for stratified randomization based on a single-factor, service network (a larger administrative unit in which the clinics were further clustered), that was highly correlated with our other potential confounding variables. This might have still enabled us to stagger randomization over time (e.g., by randomizing service networks at staggered intervals). However all of the clinic leaders were very anxious to have their clinics randomized to the intervention status due to external incentives for increasing CRC screening rates. They also wanted their clinics to be randomized as early in the year as possible to give them the maximum time to work with their populations should they be selected for the intervention. So we accepted their assurances that they could start up quickly. We opted for a very

transparent process in which the randomization was done by a state senator as part of a webcast, and all of the assignments were announced at the same time. In retrospect we should have been more skeptical of the clinics' promises and should have insisted on a staggered randomization process. But it isn't clear that the clinics would have accepted this. We had no power to compel clinics to bend to our will or even to participate in the study. Instead, we had to cooperate with them in a partnership to work out conditions for participation and implementation details.

The use of a real-time intervention tool and the ACA rollout were issues over which we had no control, and these just exacerbated the asynchrony between measurement window and intervention delivery created by the delayed start-up.

We hope that this case study of the problems we encountered and how we have chosen to deal with them might help other investigators facing similar challenges to think about such issues and the pros and cons of various strategies for dealing with them. In writing up the results of such studies, it is important to be as transparent as possible regarding the problems encountered and how they were addressed. It is also important to present one's intended primary analysis along with the alternative analyses, to make one's case in the discussion section for how the findings should be interpreted, and to let the readers draw their own conclusions.

We also caution against deciding on what secondary analyses to do after the data is in hand. In such cases it is very difficult to be intellectually honest with oneself as to which approaches to retain or reject. In our own case, we considered a number of possible analyses, in addition to those presented here, prior to collecting our outcome data, and we tried to weigh their pros and cons on a purely theoretical basis before deciding on the analyses to present.

References

1. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis* 1967 Aug; 20(8):637-648.
2. Roland M, Torgerson DJ. What are pragmatic trials? *BMJ* 1998 Jan 24; 316(7127):285.
3. Caro JJ, Ishak KJ. Optimizing the design of pragmatic trials: key issues remain. *J Comp Eff Res* 2012 Jul; 1(4):319-327.
4. Foster N, Little P. Methodological issues in pragmatic trials of complex interventions in primary care. *Br J Gen Pract* 2012 Jan; 62(594):10-11.
5. Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, Lam M, Seguin R. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol* 2003 Dec 22; 3:28.:28.
6. Murray DM. *Design and Analysis of Group-Randomized Trials*. New York, NY: Oxford University Press; 1998.
7. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol* 2006 Nov 8; 6:54.:54.
8. Coronado GD, Vollmer WM, Petrik A, Taplin SH, Burdick TE, Meenan RT, Green BB. Strategies and Opportunities to STOP Colon Cancer in Priority Populations: design of a cluster-randomized pragmatic trial. *Contemp Clin Trials* 2014 Jul; 38(2):344-349.
9. U.S.Preventive Services Task Force. Final Recommendation Statement: Colorectal Cancer: Screening. Available at: <http://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/colorectal-cancer-screening>; October 2014.
10. Centers for Medicare & Medicaid Innovation. Comprehensive Primary Care Initiative eCQM User Manual. Available at: https://www.cms.gov/eHealth/downloads/PY_2014_CPC_EHR_CQM_Manual_AUGUST.PDF; 3.0. Accessed on: 8-13-2015.
11. Johnson KE, Tachibana C, Coronado GD, Dember LM, Glasgow RE, Huang SS, Martin PJ, Richards J, Rosenthal G, Septimus E, Simon GE, Solberg L, Suls J, Thompson E, Larson EB. A guide to research partnerships for pragmatic clinical trials. *BMJ* 2014 Dec 1; 349:g6826. doi: 10.1136/bmj.g6826.:g6826.
12. Wang CY, de Dieu TJ, Anderson ML, Vernon SW, Chubak J, Fuller S, Green BB. Time to screening in the systems of support to increase colorectal cancer screening trial. *Cancer Epidemiol Biomarkers Prev* 2014 Aug; 23(8):1683-1688.
13. Mosen DM, Feldstein AC, Perrin N, Rosales AG, Smith DH, Liles EG, Schneider JL, Lafata JE, Myers RE, Kositch M, Hickey T, Glasgow RE. Automated telephone calls improved completion of fecal occult blood testing. *Med Care* 2010 Jul; 48(7):604-610.
14. National Committee for Quality Assurance (NCQA). HEDIS & Quality Compass. Available at: <http://www.ncqa.org/HEDIXQualityMeasurement/WhatIsHEDIS.aspx>. Accessed on: 5-28-2015.