



Published in final edited form as:

Contemp Clin Trials. 2025 February ; 149: 107784. doi:10.1016/j.cct.2024.107784.

Improving Colonoscopy Quality in the National VA Healthcare System

Andrew Gawron^{1,2}, Travis Bailey¹, Rachel Codden², Jason Dominitz³, Samir Gupta^{4,5}, Christian Helfrich³, Charles Kahi^{6,7}, Lila Krop^{8,9}, Carmel Malvar^{8,9}, Grace McKee^{9,10}, Morgan Millar², Ashley Mog³, Tiffany Nguyen-Vu^{8,9}, Olga Patterson², Angela P. Presson², Sameer Saini^{11,12}, Mary Whooley^{8,10}, Yiwen Yao^{1,2}, Susan Zickmund², Tonya Kaltenbach^{8,9,10}

¹VA Salt Lake City Health Care System

²University of Utah, Salt Lake City, UT

³VA Puget Sound Health Care System, Seattle, Washington

⁴VA San Diego

⁵University of California, San Diego.

⁶Indiana University School of Medicine

⁷Richard L. Roudebush Veterans Administration Medical Center

⁸San Francisco VA Medical Center

⁹Department of Medicine, University of California, San Francisco, CA

¹⁰Measurement Science Quality Enhancement Research Initiative, San Francisco VA Healthcare System

¹¹VA Ann Arbor HSR&D Center for Clinical Management Research

¹²University of Michigan School of Medicine

Abstract

Background: Colorectal cancer (CRC) prevention is a Veterans Affairs (VA) priority. Colonoscopy quality, especially adenoma detection rate (ADR), is critical for effective screening. Our research indicates considerable variation in ADR among VA providers. Even a slight increase in ADR can reduce fatal CRC rates, and audit and feedback strategies have improved ADR in other settings. A recent report identified deficiencies in VA colonoscopy quality, highlighting the need for standardized documentation and reporting. To address this, we developed the VA Endoscopy Quality Improvement Program (VA-EQuIP), which aims to improve colonoscopy quality through benchmarking and collaborative learning, aligning with VA's modernization priorities and HSR&D and QUERI goals of accelerating evidence-based implementation.

Methods: We will conduct a stepped wedge cluster randomized trial to evaluate whether VA-EQuIP improves provider ADR compared to usual care, the implementation of VA-EQuIP, site-level factors associated with colonoscopy quality improvement, and components of provider behavior change. Using mixed methods our study will measure outcomes like reach, implementation, adoption, maintenance of VA-EQuIP, and provider behavior change. The analysis will include primary and secondary outcomes, such as overall and screening ADR, cecal intubation rate, and bowel preparation quality, using mixed effects generalized linear models and interrupted time-series analyses. Adoption and implementation will be evaluated through usage statistics, surveys, and qualitative interviews to identify factors influencing success.

Discussion: This study will assess the impact of VA-EQuIP on colonoscopy quality metrics and factors associated with effective implementation. VA-EQuIP infrastructure allows for national-scale implementation and evaluation of quality reporting with minimal manual labor, guiding future quality improvement efforts to ensure optimal patient care.

Background

Colorectal cancer (CRC) prevention is a Veterans Affairs (VA) priority. Veterans are highly impacted by CRC, the third most common cancer diagnosed^{1,2} with a 35% 3-year mortality rate.^{3,4} In VA, >200,000 colonoscopies are performed annually, approximately half of which are for screening.⁵ Colonoscopy quality benchmarks have been strongly linked to CRC incidence and death.^{6,7} Adenoma detection rate (ADR), the number of patients screened who have at least one adenomatous polyp detected, has been the strongest surrogate exam quality. Our team confirmed that colonoscopy quality varies across VA, with provider ADRs ranging from 13–79%. We found that as little as a 5% increase in ADR was associated with a 4% relative decrease in fatal CRC after a normal colonoscopy.

Audit and feedback is an evidenced-based strategy shown to reduce CRC incidence and deaths through direct improvements in endoscopist performance. A previous study showed that with annual feedback and quality benchmark indicators on screening colonoscopy performance, the majority of providers (74.5%) increased their ADR.⁸ When compared with no increase in ADR, reaching or maintaining the highest quintile ADR resulted in significant decreased risk for post colonoscopy interval and fatal CRC.

A recent Office of the Inspector General (OIG) report highlighted colonoscopy quality deficiencies in VA.⁹ The report strongly recommended the requirement of standardized documentation of colonoscopy quality indicators based on professional society guidelines and published literature.⁹ In a national VA needs assessment, we identified significant manual workforce and time required to collect and report colonoscopy quality metrics and 40% of sites not even measuring ADR.¹⁰ Without a centralized, automated mechanism to continually measure or report quality, VA cannot ensure Veterans receive high quality colonoscopy.

We developed a novel informatics infrastructure for centralized colonoscopy quality reporting across VA, enabling implementation of the VA Endoscopy Quality Improvement Program (VA-EQuIP) to directly address the OIG recommendations and VA's critical need for evidence-based colonoscopy quality measurement and reporting. The National

GI Program Office will implement VA-EQuIP to provide VA sites and endoscopists with bi-annual audit and feedback of colonoscopy quality, individual provider benchmarking to local and national performance, and collaborative learning sessions moderated by national experts in colonoscopy training and quality. Learning collaboratives were originally developed by the Institute of Medicine to support peer-to-peer and peer-to-expert learning to improve healthcare.¹¹ Specifically, quality improvement learning collaboratives allow sharing of expertise across practice sites to hasten the diffusion of evidence based practices.¹² Our VA-EQuIP collaborative will emphasize shared learning across multiple units led by nationally recognized experts in colonoscopy quality. Learning sessions will include didactic sessions, discussion, and skill building activities followed by work on quality improvement projects.¹² Despite the widespread use and description of learning collaboratives, there have been very few randomized controlled trials performed with data on their effectiveness in clinical care.^{13–15} Recently, “virtual” collaboratives have been described and used in many patient safety projects (fall prevention, pressure ulcer prevention, catheter infections) in the VA.^{16–20} Virtual collaborative learning has not been evaluated as an approach to improve colonoscopy quality.

The implementation of VA-EQuIP is a rare opportunity to study the deployment and impact of a large-scale learning health system initiative.

Methods

Our goal is to evaluate, among VA providers, how the implementation of the VA Endoscopy Quality Improvement Program (VA-EQuIP) compared to usual care (C) affects the adenoma detection rate (ADR).

Aims and Objectives

Aim 1: Determine if VA-EQuIP implementation increases provider ADR compared to usual care.

Aim 2a: Evaluate VA-EQuIP implementation and identify site-level factors associated with colonoscopy quality improvement.

Aim 2b: Explore and identify components of provider behavior change after VA-EQuIP implementation.

Study Design

We will conduct a prospective, multi-center, stepped wedge cluster randomized trial design to facilitate a graduated rollout of VA-EQuIP across participating sites, which is more feasible than a parallel randomized trial.²¹ We will stratify randomization of sites to our six enrollment waves by site size and site annual ADRs in 2018. Low volume sites (<25 colonoscopies/year) and low ADR sites (<30% ADR) are likely to benefit the most from VA-EQuIP. Stratifying by these covariates helps ensure that we can adjust for facility-level confounders and account for secular trends.

Using a mixed methods approach, we will identify quantitative outcomes (e.g. reach, implementation, adoption, maintenance), qualitative measures (barriers, facilitators), and components of provider behavior change associated with VA-EQuIP implementation and colonoscopy quality improvement.

Our conceptual model (Figure 1) incorporates the Capability, Opportunity, Motivation (COM-B) model of behavior change²² and the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework²³ with integrated Qualitative Evaluation for Systematic Evaluation (RE-AIM QuEST²⁴). The evaluation targets constructs defined by REAIM-QuEST²⁴ as they apply to colonoscopy quality measurement, reporting, and improvement.

Quantitative Survey Design and Qualitative Interview Guide: Surveys will be conducted pre and post (12 months) VA-EQuIP implementation and interviews will be held after each randomized cluster. Facility steward surveys will assess site characteristics and colonoscopy protocols before and after VA-EQuIP implementation to assess site level changes. Practice changes may include changed clinical protocols or practice, such as quality measurement practices, quality reporting practices, bowel preparation protocols, endoscopy technique, use of assist devices marketed to improve adenoma detection, or seeking additional training or proctoring. The provider survey was modified from a previously developed, validated COM-B-based survey²⁵ to capture behavior change domains consistent with our conceptual model. Qualitative interviews with stewards will explore barriers and facilitators of VA-EQuIP implementation (RE-AIM QuEST).²⁵ Qualitative interviews with providers will explore components of behavior change related to VA-EQuIP implementation.

Quantitative Survey Administration: The quantitative surveys will be delivered to the identified facility stewards (N=73) and providers (N=657) for each enrolled site via the Research Electronic Data Capture (REDCap)²⁶ platform at baseline and 12 months post-intervention. We will incorporate established methods of maximizing web survey responses.²⁷

Qualitative Interviews and Transcription: Trained interviewers will conduct and record up to 60-minute telephone interviews with individual endoscopists at participating facilities. All interviews will be transcribed verbatim by the VA Health Services Research and Development Centralized Transcription Services Program. Purposive sampling²⁸ creates the sample needed for interviewing endoscopists. Based on principles of thematic saturation, we estimate a minimum sample size of 50 persons (25 facility stewards and 25 individual providers), with plans to evenly distribute interviews between high and low performing providers and at least 2 sites per VISN. Given the two cohorts of providers at high and low performing sites geographically dispersed across the country, our sample size was chosen to ensure saturation.²⁹ We will continue with interviews until thematic saturation is reached.²⁹

Eligibility:

Inclusion criteria: Sites could be included in the study if we could identify colonoscopy procedure and pathology notes in the VA central data warehouse (CDW) database for providers at the site. Of the 151 VA sites, we could identify 73 sites with reliable colonoscopy and pathology data from the CDW.

Exclusion criteria: VA facilities without existing colonoscopy procedure or pathology notes in our operational database will not be included.

Outcomes: The primary outcomes are overall and screening ADR for endoscopists after implementation of the quality report cards, with an average follow up of 19.5 months (range 12–27 months). Overall ADR is the percentage of patients undergoing colonoscopy for any indication with one or more adenomas detected. Screening ADR is the percentage of patients ≥ 50 years undergoing screening colonoscopy who have one or more adenomas detected. ADR calculations will include confidence intervals. Secondary outcomes measures include cecal intubation rate and bowel preparation quality. Cecal intubation is passage of the colonoscopy tip to a point proximal to the ileocecal valve so the entire cecum is visible. The cecal intubation rate is the proportion of colonoscopies with cecal intubation documented in the procedure note. Bowel preparation quality is determined by the endoscopist for bowel cleanliness allowing polyp visualization and removal.

Implementation strategy: VA-EQuIP will be implemented under the National GI Program Office with two primary implementation strategies: audit and feedback and collaborative learning. A facility steward will serve as the point person for communications with the VA-EQuIP team and lead local quality improvement activities (Table 1 and Figure 2).

Operational Database: We created an operational database of colonoscopy procedures and linked pathology notes housed in the VA CDW Text Integration Utility that can be semi-automatically updated with the most current data across the VA healthcare system. We optimized a machine learning document classifier that delineates a note as either a colonoscopy procedure note or not. This allows us to rapidly and prospectively confirm colonoscopy notes for NLP processing and colonoscopy quality reporting with excellent performance (Table 2).

The VA-EQuIP dashboard presents biannual and cumulative colonoscopy quality metrics to VA sites and providers and provides comparison metric data at the site and national level (Figure 3).

Data collection

CDW structured data include patient demographics and endoscopist name and specialty. Variables derived from NLP of colonoscopy procedure and pathology notes include extent of exam, screening indication, and adenoma detection. We will use validated algorithms to ascertain colorectal adenoma and adenocarcinoma for adenoma detection.^{30–32}

Analysis

We will compare the efficacy of VA-EQuIP versus usual care for the primary outcomes of overall and screening indication ADR and secondary outcomes of bowel preparation quality and cecal intubation rate. We hypothesize that VA-EQuIP will lead to a higher ADR, bowel preparation quality, and cecal intubation relative to standard of care, and the effect will be particularly pronounced among providers who have low quality metrics prior to our intervention. The primary analysis takes advantage of the stepped wedge design by simultaneously comparing pre- and post- intervention data both within and across providers to evaluate the effect of the intervention while controlling for secular trends, which are uniform across sites. As a secondary analysis, we will implement a two-stage interrupted time-series and meta-regression analysis. For all analyses, our primary inferences evaluate the ratio of ADR with versus without intervention for a given provider. However, our planned analyses also address the data structure, which nests patients within providers within sites.

Primary Analysis: We will use patient data at the provider level at monthly intervals to examine ADR changes pre- and post-intervention. Data collection spans 30 months, from January 1, 2020, to June 30, 2022, including baseline data from 2019 to facilitate an immediate intervention start. ADR rates for 2018 will determine randomization strata (high vs. low ADR). Our stepped wedge design yields varying pre- and post-intervention periods for each enrollment wave or “cluster” (Figure 5). New sites will be enrolled quarterly in 6 waves (12–13 wave, 73 total), with data from the 3-month intervention implementation period being discarded. We will augment baseline data capture by 12 months and extend data capture following our final wave by another 12 months to improve estimation of pre/post intervention effects within individual providers in the first and final waves. Pre- and post-intervention data will range from 12–27 months depending on the wave, averaging 19.5 months. All providers, both part-time and full-time will be captured. Our primary intention to treat (ITT) analysis required each provider to perform at least one procedure during the pre-intervention time period. We conducted the following per protocol analyses: 1) each provider performed at least 5 procedures during the pre-intervention time period and at least 5 procedures in the post-intervention time period, 2) we additionally required provider’s data to be displayed in the dashboard, and 3) each provider performed at least 25 procedures during the pre-intervention time period and at least 25 procedures during the post-intervention time period. We will also capture colonoscopy procedure volume, and include confidence intervals for the ADR calculation.

We will construct a mixed-effects generalized linear model that will act on data aggregated monthly within each provider to estimate the average intervention effect across providers. A Poisson outcome model with a log link will characterize each provider’s monthly rate, with the number of positive screens as the outcome and the log of total screens as an offset term.^{33,34} In the case of overdispersion, we will use a negative binomial outcome model. The main predictor of interest will be a pre/post indicator for whether the report card intervention was in effect at a particular month. We will control for time, site characteristics (urban/rural, cumulative procedure volume in 2018, current volume, site ADRs in 2018, use of split dose bowel preparation protocols in 2018, and whether or not ADR was measured

in 2018), provider characteristics (years in practice, practice type, procedure volume), and baseline patient characteristics (age, sex, race/ethnicity). Study time will be coded in quarters, to incorporate seasonal effects in addition to secular trends over the study period. Random effects will be included for both provider and site, where provider is nested within site. Let $PositiveScreens_{ijt}$ denote the observed number of positive screens for the j th provider at site i during month t , and let $TotalScreens_{ijt}$ denote the total number of patients screened. Because the monthly screening rates are expected to be relatively low (< 0.30), we assume that the observed number of positive screens, $PositiveScreens_{ijt}$ follows a Poisson distribution with conditional mean λ_{ijt} . We anticipate the following model for λ_{ijt} on a log scale, where $TotalScreens_{ijt}$ is included as an offset term on the right hand side of the equation:

$$\log(\lambda_{ijt}) = \beta_0 + \log(TotalScreens_{ijt}) + \beta_1 Intervention_{ijt} + \beta_2 Time_{ijt} + \beta_3 Urban_i + \beta_4 SiteVolume_i + \beta_5 YrsPractice_{ij} + \beta_6 PracticeType_{ij} + \beta_7 ProviderVolume_{ij} + \beta_8 Age_{ij} + \beta_9 Sex_{ij} + \beta_{10} Race_{ij} + \beta_{11} Volume2018_i + \beta_{12} ADR2018_i + \beta_{13} SDBP2018_i + \beta_{14} ADRMeasured_i + \alpha_i + \mu_{j(i)}$$

where β_0 is the intercept, β_2 is a vector of coefficients for the study quarter as expressed by $Time_{ijt}$, α_i indicates the random effect for site i , $\mu_{j(i)}$ indicates the random effect for provider j within site i . $Volume2018_i$ (high/low), $ADR2018_i$ (high/low) and $ADRMeasured_i$ (yes/no) are the indicator variables used in stratifying the baseline randomization of sites to waves. We will also examine whether the intervention effect varies by time since intervention and by provider characteristics by considering these interaction terms in the model. Under the proposed Poisson model and logarithmic link function, exponentiating the results will allow us to report rate ratios (RRs), their 95% confidence intervals (CIs), and p-values from the models. Under the generalized linear mixed model the estimated treatment effect, obtained by exponentiating the estimate of β_1 , represents a rate ratio (RR) corresponding to the screening rate after versus before implementing the intervention for a given provider.

We will plot monthly ADRs for a random sample of endoscopists to examine ADR rate changes over time and check for a lagged intervention effect. Anticipating a one-month lag, we will exclude outcome data for this period to avoid biased estimates. We will examine correlation between successive monthly ADR rates within providers using residual plots and the Durbin-Watson or Durbin's alternative test and include study time lags as appropriate.³⁵ If analyses suggest that the residual covariance deviates substantially from those assumed by the nested random effects model, we will substitute a generalized estimating equation (GEE) analysis, assuming outcomes are independent between sites, and use robust standard errors for statistical inference.³⁶ The GEE approach can also be used as a fallback if the nested random effect structure proves to be numerically intractable.

As a secondary question we will assess the impact of compliance with the intervention on ADR rate among the providers who responded to our survey. Compliance will be measured by dashboard access and participation in collaborative learning sessions. Compliance will be modeled as an interaction with intervention in the model described above, to see if the efficacy of the intervention varies by compliance status.

Power Calculation: Our power calculation for the intervention effect comes from a stepped wedge model. In 2017, 73 sites with 657 providers performed 155,926 cases per year in our sampling frame. We estimate a 5% loss of site data and an 8% provider attrition

rate based on data provided by the National GI Program Office⁵, leaving 69 sites with 571 providers seeing 135,517 patients/year, or 95 providers/wave and 59 patients/provider/quarter. The number of patients seen by providers factors into our power calculation as it affects the standard errors of the intervention effect, where seeing more patients yields greater power. Our primary outcome, ADR, is conservatively estimated to range from 13–79% based on prior research^{6,8} and our preliminary data. For cluster randomized trials using implementation designs such as ours, the intra-cluster correlation coefficient (ICC) estimates for patient outcomes are approximately 0.030.³⁷ Based on an ADR of 27.5% in the pre-intervention period, we have about 93% power to detect a 1.0% absolute increase in ADR. This difference is clinically important, because a 1% increase in ADR is associated with a 3.0% decrease in risk of incident CRC, and a 5% decrease in risk of fatal interval CRC.⁶ Power calculations were conducted in R using the `swCRTdesign` package, and are conservative due to exclusion of three-quarters of baseline data and because the second level of clustering by provider is not considered.^{21,38}

Secondary Analysis: While the analysis described above will be considered primary, our stepped wedge design coupled with our focus on provider performance provides opportunity for several analysis options. To report on the intervention effect efficacy to individual providers, we will conduct an interrupted time-series (ITS) analysis for each provider. As a secondary analysis, we will then estimate the average intervention effect across providers by using a meta-regression approach to combine results across providers.^{39,40} This analysis is conventional for interventions implemented at different time points for different units,^{40,41} and has the advantage of evaluating both the overall intervention effect and the heterogeneity in the treatment effect across sites and providers.

The first stage ITS analysis will consist of segmented regression with an autoregressive error model to account for the correlation between successive monthly ADRs. We assume that a provider's observed number of positive screens during month t , $PositiveScreens_t$, follows a Poisson distribution with conditional mean λ_t .

Our basic ITS model for λ_t for a particular provider has the following form:

$$\text{Log}(PositiveScreens_t) = \beta_0 + \text{log}(TotalScreens_t) + \beta_1 Intervention_t + \beta_2 Time_t + \beta_2 Time - After_t - Intervention_t + \beta_3 Age + \beta_3 Sex + \beta_{10} Race + \beta_{11} Season_t$$

where t indicates time coded in months.⁴² In this analysis, time is coded monthly and treated as a continuous variable. The main predictors of interest will include the immediate intervention effect ($Intervention$) and whether the monthly post-intervention time trend differs from the pre-intervention time trend ($Time - After - Intervention$). Autocorrelation will be examined using residual plots and the Durbin-Watson or Durbin's alternative test and corrected using lagged study time variables as appropriate.³⁵ Each model will provide a RR estimate, 95% CI, and p-value for a provider's intervention effect.

The second stage analysis will consist of random effects meta-regression to aggregate the provider intervention effects estimated from the first stage.^{39,40} This will estimate the average intervention effect across providers, accounting for heterogeneity at the site and provider levels by including random effects for both. We include a random effect for provider because we conservatively consider the providers included in our analysis to be a random sample from the full possible set of providers.⁴⁴ Konstantopoulos et al.

describe a similar multilevel meta-regression analysis where studies are nested within school districts.⁴⁴ Our meta-regression model will adjust for randomization stratification procedure variables (*Volume2018*, *ADR2018*, and *ADRMasured*), provider-level variables (baseline years in practice and annual procedure volume), and site level characteristics (split dose bowel preparation protocols (SDBP) in 2018, urban/rural, and cumulative annual procedure volume).

Power Calculation: Our power calculation for the overall intervention effect is estimated from an ITS model, accounting for 5% data loss at the site level and 8% provider attrition. Each provider has an average of 19.5 months of data pre- and post-intervention, totaling an estimation of 386 patients/provider or 220,215 patients per period. The number of patients seen by providers greatly affects the power of our design, so we include it in our calculation. We simulated ITS data with varying intervention effect sizes adapting code provided by Rozario et al.⁴⁵ Under an ITS model with an autocorrelated error structure, we expect to achieve 96.8% power to detect a 1.25% increase in ADR from a pre-intervention rate of 27.5% (based on 1000 simulations). Power calculations were conducted in SAS v 9.0.

Explanatory Analysis: One potential limitation of the analyses described above is aggregating patient-level data to the provider month-level, which is susceptible to Simpson's paradox, where the direction of a relationship between an exposure and outcome can change when the data is analyzed at an aggregated level versus a partitioned level.^{46,47} The ITS analysis, comparing pre/post within providers rather than intervention and control at the same time, is particularly vulnerable. To overcome this, we will re-examine both our primary stepped wedge analysis and secondary ITS analysis using patient-level data. We will again use Poisson outcome models to report RRs (but without an offset term). The interpretation of the results will be the same as before. For the stepped wedge analysis, the exponentiated intervention coefficient will represent the ratio of screening rates post-intervention versus pre-intervention for a given provider. For the ITS analysis, the first stage modeling will yield the ratio of each provider's screening rate post- versus pre-intervention. The second stage will yield a pooled estimate of the intervention RR across providers.

VA-EQuIP Statistics: To evaluate VA-EQuIP adoption and implementation we will continuously monitor the number of providers viewing dashboards and participating in collaborative learning sessions through 12-month assessment time points. Descriptive statistics will be performed at the site and provider level.

Survey analysis: Survey data analysis will include descriptive statistics of facility steward and individual provider survey responses at baseline and 12 months post-implementation. This includes response distributions or means and standard deviations for all baseline site characteristics and implementation and behavior change measures. We will examine the percentage of providers who viewed their quality data and whether site-level activities impacting provider behavior ("opportunities") were initiated post-intervention. The facility steward survey will assess site protocol changes and the sustainability of new practices initiated earlier in the intervention. Provider survey responses will be summarized using the adapted COM-B survey scoring and compared between baseline and 12 months. We will

also compare COM-B measures between providers, stratified by ADR categories pre- and post-VA-EQuIP, and by involvement in collaborative learning as a categorical variable.

Qualitative analysis: Verbatim transcripts from qualitative interviews will be analyzed using the qualitative “editing” method.⁴⁸ Representative quotations will be captured using Atlas.ti (Scientific Software, Berlin Germany) and mapped to COM-B behavior change domains. With Dr. Zickmund, two trained qualitative analysts will meet and process any differences until they agree. Agreed codes will be recorded in a master file for the final analysis. Independent coding will ensure narrative coherence and inter-coder reliability. Inter-coder reliability kappa scores of below 0.61 will trigger a training meeting. This process consistently achieves an inter-coder reliability kappa score of 0.70.⁴⁹ After coding, analysts synthesized the perceived behavior changes, barriers, facilitators, and recommendations across participants into a thematic report that was used to more deeply understand site-level context.

Evaluation of VA-EQuIP implementation: We will use an explanatory mixed-methods approach for evaluation, consisting of three components: monitoring VA-EQuIP usage statistics, pre/post surveys of facility stewards and providers, and qualitative interviews with facility stewards and individual providers. The approach will capture key RE-AIM (2a) and COM-B (2b) domains with qualitative analysis to more fully explain barriers and facilitators to implementation and components of provider behavior change.

Aim 2a: Evaluation measures will include reach, adoption, implementation, and maintenance domains (Appendix 5). Reach will be measured using baseline facility steward survey site practice characteristics and claims data from CDW fee basis, VA Choice, and MISSION Acts, and contracting records to determine the proportion of Veterans having colonoscopies performed by VA versus non-VA providers.

Our comprehensive data infrastructure allows electronic monitoring of VA-EQuIP quality data views by site and provider. These data will serve as quantitative, continuous measures of adoption at site and individual provider levels. We will also collect participant data from sites and providers participating in collaborative learning sessions and include additional adoption measures in the surveys, asking about use of VA-EQuIP for quality measurement and reporting at each site. For implementation measures, the centralized VA-EQuIP infrastructure will capture fidelity by assessing the timing of intervention delivery. Facility steward surveys will capture outcome measures for sites before and after VA-EQuIP implementation (Appendices 6–7). Qualitative interviews with facility stewards will explore more in depth site-level barriers and facilitators to VA-EQuIP implementation and start after the last collaborative learning session for each randomized cluster.

Aim 2b: Provider surveys and qualitative interviews were designed to capture components of behavior change mapped COM-B model domains before and after VA-EQuIP implementation (Appendices 6–7). Provider qualitative interviews will start after the last collaborative learning session for each randomized cluster.

Discussion

Colonoscopy quality is linked to patient outcomes. The national VA healthcare system lacks a comprehensive program to measure and report colonoscopy quality and help providers improve adenoma detection.⁹ Accurate measurement of quality metrics is challenging because validated quality metrics are not available in structured VA data from VA CDW. Colonoscopy procedure documentation resides in text notes in Vista/CPRS or endoscopic reporting software. Commonly used VA endoscopy note writer software programs do not facilitate tracking of pathology data and quality measurement. These same issues will persist unless infrastructure is in place to collect and process colonoscopy quality data. Thus, within VA, there has not been a reliable, efficient way to measure colonoscopy quality and ensure optimal protection from CRC incidence and death for Veterans.

VA-EQuIP has a high probability of improving a quality metric outcome (ADR) directly associated with CRC death, a common cancer in Veterans. This randomized controlled trial presents a unique opportunity to demonstrate that even small improvements in clinical performance through audit and feedback, coupled with skill and quality improvement support, can save lives. A 1% increase in ADR can lead to a 3% reduction in interval cancer death. Our randomized program evaluation is a tremendous opportunity to determine the large-scale effect of the VA-EQuIP strategy on changes in the colonoscopy quality metrics of individual endoscopists over time. The evaluation will identify factors associated with effective implementation and colonoscopy quality improvement. Our prior and planned work builds toward our long-term goal to reduce mortality in Veterans by enhancing early CRC detection and inform national quality improvement initiatives, including remediation training for consistently low-performing endoscopists.

The novel foundation of VA-EQuIP is an NLP algorithm and informatics reporting infrastructure we developed to measure and report the quality of colonoscopies performed by individual VA colonoscopists.¹⁰ Our program is similar to the VA Surgical Quality Improvement Program (VASQIP),^{50,51} however, VASQIP requires significant manual labor with a dedicated nurse manager at every VA site to collect and enter data. Our novel informatics infrastructure collects and enters the data into the colonoscopy registry. This automation allows implementation and evaluation of quality reporting and improvement on a national scale in a shorter time frame with minimal manual labor.

References:

1. Zullig LL, Smith VA, Jackson GL, et al. Colorectal Cancer Statistics From the Veterans Affairs Central Cancer Registry. *Clin Colorectal Cancer*. 2016;15(4):e199–e204. doi:10.1016/j.clcc.2016.04.005 [PubMed: 27301717]
2. Zullig LL, Jackson GL, Dorn RA, et al. Cancer incidence among patients of the U.S. Veterans Affairs Health Care System. *Mil Med*. 2012;177(6):693–701. doi:10.7205/milmed-d-11-00434 [PubMed: 22730846]
3. Brenner H, Stock C, Hoffmeister M. Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. *BMJ*. 2014;348:g2467. doi:10.1136/bmj.g2467 [PubMed: 24922745]

4. Meester RGS, Doubeni CA, Lansdorp-Vogelaar I, et al. Colorectal cancer deaths attributable to nonuse of screening in the United States. *Ann Epidemiol.* 2015;25(3):208–213.e1. doi:10.1016/j.annepidem.2014.11.011 [PubMed: 25721748]
5. Kaltenbach T. Cyberseminar Transcript Series: Using Data and Information Systems in Partnered Research Session: Using Data and Information Systems to Measure Colonoscopy Quality, , July 18, 2017 https://www.hsr.dva.gov/for_researchers/cyber_seminars/catalog/transcripts/2330.doc
6. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med.* 2014;370(14):1298–1306. doi:10.1056/NEJMoa1309086 [PubMed: 24693890]
7. Kaminski MF, Regula J, Kraszewska E, et al. Quality indicators for colonoscopy and the risk of interval cancer. *N Engl J Med.* 2010;362(19):1795–1803. doi:10.1056/NEJMoa0907667 [PubMed: 20463339]
8. Kaminski MF, Wieszczy P, Rupinski M, et al. Increased Rate of Adenoma Detection Associates With Reduced Risk of Colorectal Cancer and Death. *Gastroenterology.* 2017;153(1):98–105. doi:10.1053/j.gastro.2017.04.006 [PubMed: 28428142]
9. Alleged Access Delays and Surgery Service Concerns, VA Roseburg Healthcare System. Office of Healthcare Inspections.
10. Gawron AJ, Lawrence P, Millar MM, et al. A Nationwide Survey and Needs Assessment of Colonoscopy Quality Assurance Programs in the VA. *Fed Pract Health Care Prof VA DoD PHS.* 2018;35(3):26–32.
11. Egberg MD, Gulati AS, Gellad ZF, Melmed GY, Kappelman MD. Improving Quality in the Care of Patients with Inflammatory Bowel Diseases. *Inflamm Bowel Dis.* 2018;24(8):1660–1669. doi:10.1093/ibd/izy030 [PubMed: 29718299]
12. Bunker AC, Hanson RF, Doogan NJ, Powell BJ, Cao Y, Dunn J. Can Learning Collaboratives Support Implementation by Rewiring Professional Networks? *Adm Policy Ment Health.* 2016;43(1):79–92. doi:10.1007/s10488-014-0621-x [PubMed: 25542237]
13. Nadeem E, Olin SS, Hill LC, Hoagwood KE, Horwitz SM. A literature review of learning collaboratives in mental health care: used but untested. *Psychiatr Serv Wash DC.* 2014;65(9):1088–1099. doi:10.1176/appi.ps.201300229
14. Nadeem E, Olin SS, Hill LC, Hoagwood KE, Horwitz SM. Understanding the components of quality improvement collaboratives: a systematic literature review. *Milbank Q.* 2013;91(2):354–394. doi:10.1111/milq.12016 [PubMed: 23758514]
15. Schouten LMT, Hulscher MEJL, van Everdingen JJE, Huijsman R, Grol RPTM. Evidence for the impact of quality improvement collaboratives: systematic review. *BMJ.* 2008;336(7659):1491–1494. doi:10.1136/bmj.39570.749884.BE [PubMed: 18577559]
16. Zubkoff L, Neily J, Mills PD, et al. Using a virtual breakthrough series collaborative to reduce postoperative respiratory failure in 16 Veterans Health Administration hospitals. *Jt Comm J Qual Patient Saf.* 2014;40(1):11–20. doi:10.1016/s1553-7250(14)40002-3 [PubMed: 24640453]
17. Zubkoff L, Neily J, King BJ, et al. Virtual Breakthrough Series, Part 1: Preventing Catheter-Associated Urinary Tract Infection and Hospital-Acquired Pressure Ulcers in the Veterans Health Administration. *Jt Comm J Qual Patient Saf.* 2016;42(11):485–AP2. doi:10.1016/S1553-7250(16)42091-X [PubMed: 28266917]
18. Zubkoff L, Neily J, King B, et al. Preventing Pressure Ulcers in the Veterans Health Administration Using a Virtual Breakthrough Series Collaborative. *J Nurs Care Qual.* 2017;32(4):301–308. doi:10.1097/NCQ.0000000000000242 [PubMed: 27902531]
19. Zubkoff L, Neily J, Quigley P, et al. Virtual Breakthrough Series, Part 2: Improving Fall Prevention Practices in the Veterans Health Administration. *Jt Comm J Qual Patient Saf.* 2016;42(11):497–AP12. doi:10.1016/S1553-7250(16)42092-1 [PubMed: 28266918]
20. Zubkoff L, Neily J, Quigley P, et al. Preventing Falls and Fall-Related Injuries in State Veterans Homes: Virtual Breakthrough Series Collaborative. *J Nurs Care Qual.* 2018;33(4):334–340. doi:10.1097/NCQ.0000000000000309 [PubMed: 29240572]
21. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials.* 2007;28(2):182–191. doi:10.1016/j.cct.2006.05.007 [PubMed: 16829207]

22. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci IS*. 2011;6:42. doi:10.1186/1748-5908-6-42 [PubMed: 21513547]
23. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health*. 1999;89(9):1322–1327. doi:10.2105/ajph.89.9.1322 [PubMed: 10474547]
24. Forman J, Heisler M, Damschroder LJ, Kaselitz E, Kerr EA. Development and application of the RE-AIM QuEST mixed methods framework for program evaluation. *Prev Med Rep*. 2017;6:322–328. doi:10.1016/j.pmedr.2017.04.002 [PubMed: 28451518]
25. Lydon S, Grealley C, Tujjar O, et al. Psychometric evaluation of a measure of factors influencing hand hygiene behaviour to inform intervention. *J Hosp Infect*. 2019;102(4):407–412. doi:10.1016/j.jhin.2019.02.003 [PubMed: 30769147]
26. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–381. doi:10.1016/j.jbi.2008.08.010 [PubMed: 18929686]
27. Smyth J, Christian LM, Dillman D. Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method, 4th Edition | Wiley. Wiley.com. Accessed June 4, 2024. <https://www.wiley.com/en-fr/Internet%2C+Phone%2C+Mail%2C+and+Mixed-Mode+Surveys%3A+The+Tailored+Design+Method%2C+4th+Edition-p-9781118456149>
28. Coyne IT. Sampling in qualitative research. Purposeful and theoretical sampling; merging or clear boundaries? *J Adv Nurs*. 1997;26(3):623–630. doi:10.1046/j.1365-2648.1997.t01-25-00999.x [PubMed: 9378886]
29. Sandelowski M. Sample size in qualitative research. *Res Nurs Health*. 1995;18(2):179–183. doi:10.1002/nur.4770180211 [PubMed: 7899572]
30. Gawron AJ, Yao Y, Gupta S, Cole G, Whooley MA, Dominitz JA, Kaltenbach T. Simplifying Measurement of Adenoma Detection Rates for Colonoscopy. *Dig Dis Sci*. 2021 Sep;66(9):3149–3155. doi: 10.1007/s10620-020-06627-2. Epub 2020 Oct 8. [PubMed: 33029706]
31. Earles A, Liu L, Bustamante R, Coke P, Lynch J, Messer K, Martínez ME, Murphy JD, Williams CD, Fisher DA, Provenzale DT, Gawron AJ, Kaltenbach T, Gupta S. Structured Approach for Evaluating Strategies for Cancer Ascertainment Using Large-Scale Electronic Health Record Data. *JCO Clin Cancer Inform*. 2018 Dec;2:1–12. doi: 10.1200/CCI.17.00072.
32. Gawron AJ, Mckee G, Dominitz JA, Yao Y, Whooley M, Kaltenbach T. Validation of a National Pathology Database for Colonoscopy Quality Reporting and Assurance. *Clin Gastroenterol Hepatol*. 2024 Aug 28:S1542–3565(24)00787–0. doi: 10.1016/j.cgh.2024.08.017. Epub ahead of print.
33. Kleinbaum DG, Kupper LL, Nizam A, Rosenberg ES. *Applied Regression Analysis and Other Multivariable Methods*. Cengage Learning; 2013.
34. Frome EL. The analysis of rates using Poisson regression models. *Biometrics*. 1983;39(3):665–674. [PubMed: 6652201]
35. Durbin J. Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables. *Econometrica*. 1970;38(3):410–421. doi:10.2307/1909547
36. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988;44(4):1049–1060. [PubMed: 3233245]
37. Campbell MK, Fayers PM, Grimshaw JM. Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research. *Clin Trials Lond Engl*. 2005;2(2):99–107. doi:10.1191/1740774505cn071oa
38. R Foundation for Statistical Computing. *R: A language and environment for statistical computing*. Published online 2018.
39. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Stat Med*. 1995;14(4):395–411. doi:10.1002/sim.4780140406 [PubMed: 7746979]
40. Gebiski V, Ellingson K, Edwards J, Jernigan J, Kleinbaum D. Modelling interrupted time series to evaluate prevention and control of infection in healthcare. *Epidemiol Infect*. 2012;140(12):2131–2141. doi:10.1017/S0950268812000179 [PubMed: 22335933]

41. Taljaard M, McKenzie JE, Ramsay CR, Grimshaw JM. The use of segmented regression in analysing interrupted time series studies: an example in pre-hospital ambulance care. *Implement Sci IS*. 2014;9:77. doi:10.1186/1748-5908-9-77 [PubMed: 24943919]
42. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther*. 2002;27(4):299–309. doi:10.1046/j.1365-2710.2002.00430.x [PubMed: 12174032]
43. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21(11):1559–1573. doi:10.1002/sim.1187 [PubMed: 12111920]
44. Konstantopoulos S. Fixed effects and variance components estimation in three-level meta-analysis. *Res Synth Methods*. 2011;2(1):61–76. doi:10.1002/jrsm.35 [PubMed: 26061600]
45. Rozario NL. Interrupted Time Series Power Calculation using DO Loop Simulations.
46. Simpson EH. The Interpretation of Interaction in Contingency Tables. *J R Stat Soc Ser B Methodol*. 1951;13(2):238–241.
47. Fu P, Panneerselvam A, Clifford B, et al. Simpson's paradox - aggregating and partitioning populations in health disparities of lung cancer patients. *Stat Methods Med Res*. 2015;24(6):937–948. doi:10.1177/0962280211434179 [PubMed: 22246415]
48. Miller W, Crabtree B. Primary Care Research: A Multimethod Typology and Qualitative Roadmap. *Doing Qual Res Prim Care Mult Strateg*. Published online April 1, 1992. <https://scholarlyworks.lvhn.org/family-medicine/45>
49. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174. [PubMed: 843571]
50. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. *Ann Surg*. 1998;228(4):491–507. doi:10.1097/00000658-199810000-00006 [PubMed: 9790339]
51. Massarweh NN, Kaji AH, Itani KMF. Practical Guide to Surgical Data Sets: Veterans Affairs Surgical Quality Improvement Program (VASQIP). *JAMA Surg*. 2018;153(8):768–769. doi:10.1001/jamasurg.2018.0504 [PubMed: 29617543]

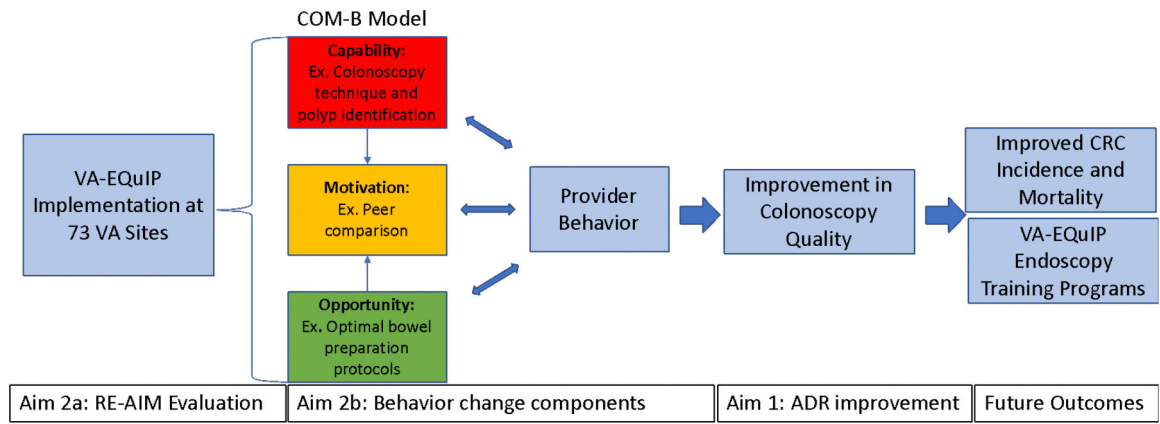


Fig. 1. Conceptual model of the VA Endoscopy Quality Improvement Program (VA-EQUIP).

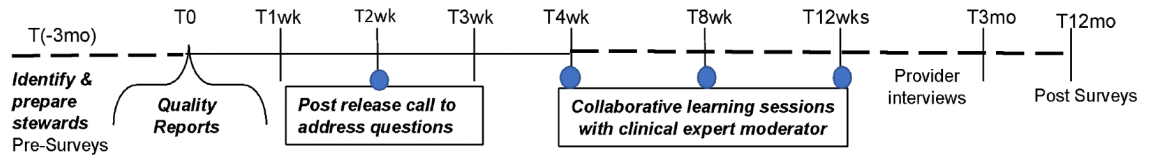


Figure 2: VA-EQuIP with strategies and evaluation timeline per randomized cluster

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

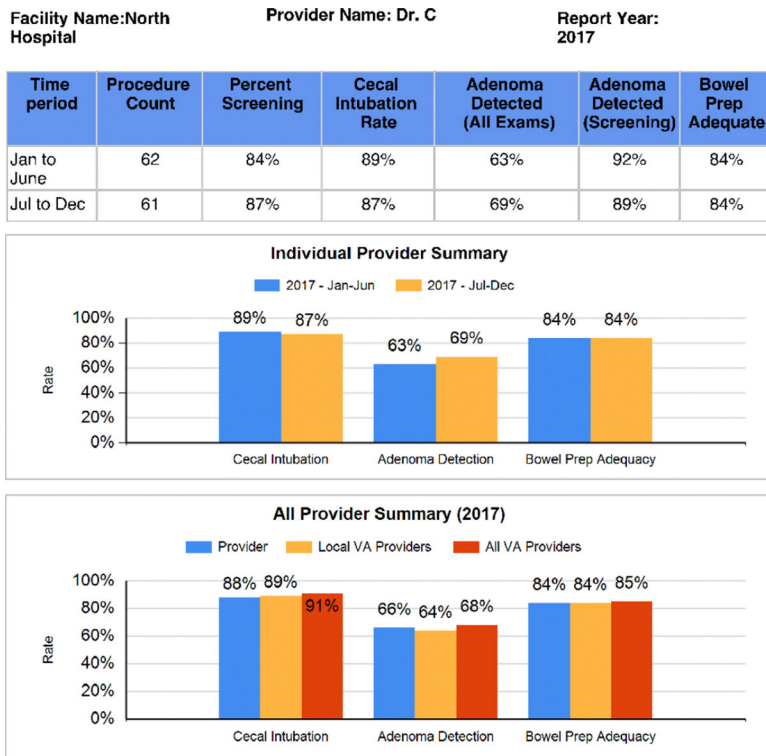


Fig. 3. The VA-EQUIP Colonoscopy Quality Provider Dashboard.

Table 1:

VA-EQuIP Implementation strategy

Implementation Strategy	Operationalization per cluster (12–13 sites)
Identify and prepare facility stewards. Facility stewards must have a VA administrative or leadership authority (i.e. GI Section Chief) to view all provider data.	T(-3mo): - GI program office will require all sites to confirm at least one facility steward responsible for colonoscopy quality assurance and site level permissions for quality data - Facility stewards will confirm with VA-EQuIP project manager colonoscopy provider names - VA-EQuIP data team will confirm that LSV data permissions for stewards and each provider are correct for quality reports o <i>Facility stewards: All site provider data access</i> o <i>Providers: Only individual data access</i> T(-1 mo): Kickoff conference call (per cluster) for stewards and providers: a) To review and demonstrate the VA-EQuIP dashboard prior to releasing VA-EQuIP reports to stewards and providers b) Promote involvement in collaborative learning sessions
Audit and provide feedback	T(0): VA-EQuIP emails with dashboard links to stewards and providers
Provide centralized technical assistance	Ongoing. Technical issues will be and triaged via trouble tickets input to the VA-EQuIP SharePoint site T(2 wks): post release conference call to address technical issues
Train and educate stakeholders • Virtual Collaborative Learning	T(0–2 wks): GI program office to send promotional announcement for collaborative learning sessions and encourage participation: T(2wks): Identify facility stewards and providers wanting to participate in collaborative learning T(2–4wks): Distribute resources / documents to collaborative learning participants: “Change package”, ⁵² Plan Do Study Act worksheets, schedule T(4–12 wks): Collaborative learning calls (Adobe Connect®) with clinical and quality improvement experts: 1) Review Change Package for evidence based practices to improve ADR: Evidence based bowel preparation protocols, Polyp identification skills, Colonoscopy technique (cleaning / washing / inspection techniques) 2) Address QI strategies for change (PDSA) 3) Troubleshooting / discussion

Table 2:

NLP Performance for Colonoscopy Quality Metrics

Variable	PPV	Sensitivity	F Measure
ADR	98%	100%	99%
Screening Indication	89%	100%	94%
Cecal Intubation Rate	98%	99%	99%
Bowel Prep Adequate	100%	100%	100%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript