Improving Colonoscopy Quality for Colorectal Cancer Screening in the National VA Healthcare System

NCT04389957

December 15, 2019

BACKGROUND

Colorectal cancer (CRC) prevention is a top VA priority. Veterans are highly impacted by CRC, the third most common cancer diagnosed[11,12] with a 3-year mortality rate of 35%.[11] Colonoscopy screening can save lives, and is the cornerstone of effective prevention. The National Polyp Study showed that removal of adenomas during screening colonoscopy decreases the subsequent development of CRC by up to 90% and death by up to 50% in comparison to historical controls. In VA, colonoscopy is the primary CRC screening modality with over 200,000 colonoscopies performed each year.[15] Eighty-two percent of Veterans are up-to-date on CRC screening, with 89% having a colonoscopy.[15] However, the quality of VA colonoscopies is unknown.

High quality colonoscopy is critical for CRC prevention. Observational studies found significant rates of CRC even after normal colonoscopy, likely due to colonoscopists' varied performance in detecting and removing polyps.[16,19] Among Medicare beneficiaries, approximately 7% of all CRC occurred in individuals who had colonoscopy that apparently missed the CRC diagnosis. A pooled analysis from eight surveillance studies suggested 52% of incident CRCs after colonoscopy were due to missed lesions.[20] These interval cancers highlight the need to focus on colonoscopy quality indicators.[16,19,21,22]

Adenoma detection rate (ADR) significantly varies by provider and has been strongly linked to both CRC incidence and mortality. In essence, polyps that are not found cannot be removed for CRC prevention. ADR, defined as the proportion of screening colonoscopies performed by a physician that has one or more histologically-confirmed adenomatous polyps or CRC, is the primary benchmark for colonoscopy inspection quality. Increasing ADR correlates with a lower risk of interval and fatal CRC.[7,16] A landmark US study of 314,872 colonoscopy exams showed marked variability in providers' ability to detect polyps, with the ADR ranging from 7.4 to 52.5%.[7] In patients of providers with the highest ADRs, as compared with those in the lowest, the adjusted hazard ratios (HR) for fatal interval CRC was 0.38 (95% CI, 0.22 to 0.65).[7] Each 1% increase in ADR was associated with a decrease risk of 3% for CRC incidence and 5% for CRC death.[7]

We have powerful pilot results showing colonoscopy quality variability in VA and direct association with CRC death.[23,24] From 1999-2011, 634,331 Veterans had a baseline colonoscopy performed by one of 3101 colonoscopists with ADRs ranging from 13-79%. In patients who had a normal colonoscopy, the higher the provider ADR, the more protected the patient was from developing future CRC: adjusted HR (adjHR) 0.57 for incident (95%CI: 0.42-0.79; ptrend< 0.001) and 0.73 for fatal (95%CI: 0.50-1.06; p=0.047) CRC for the highest vs. lowest ADR quintile. A 5% absolute increase in ADR was associated with relative reductions in fatal CRC risk of 4% after normal colonoscopy (adjHR 0.96, 95%CI: 0.95-0.97). Other important quality metrics, including bowel preparation quality[25–28] and cecal intubation rate[16] also impact patient outcomes. Both poor bowel preparation and incomplete examinations, for example, are associated with missed lesions[29] and need for earlier repeat procedures.

Interventions to improve colonoscopy quality. [The reason endoscopists perform suboptimal colonoscopies appears to be due to both issues of awareness and motivation related to their own performance, and deficits in knowledge and skills related to best practices. There is strong evidence from 3 studies that providing colonoscopy performance feedback[2,8,30] improves quality (ADR) in a reasonable time frame (3-24 mo) and patient outcomes.[2,8] Kahi and colleagues showed in a single-center VA setting that a quarterly report card improved colonoscopy quality.[2] Similarly, in a non-VA US practice of 20 endoscopists, a quality report card and implementation of practice standards resulted in a 11% overall increase in ADR.] Most recently, in a large prospective European cohort study evaluating annual feedback and quality benchmark indicators on screening colonoscopy performance,[8] the majority of the endoscopists (74.5%) increased their annual ADR. Moreover, individuals examined by endoscopists in the highest ADR quintile (> 24.6%) had significantly lower risk of interval CRC and death. When compared with no increase in ADR, reaching or maintaining the highest quintile ADR decreased adjusted hazard ratios for interval and fatal CRC to 0.27 (95% CI, 0.12-0.63; *P* = .03), and 0.18 (95% CI, 0.06-0.56; *P* = .03). Such data shows that audit and feedback alone can improve colonoscopy quality. However, it may not be enough for every provider. A Polish study found that an intensive endoscopy hands-on training session compared to feedback alone improved ADRs 4.5% (vs. 2.3% in feedback only group) at 12 months.[31] [And a recent US abstract highlighted that in addition to audit and feedback, a single online training module offered to providers improved ADR by 5.9% in 12-24 months.[32] Feedback and

benchmarks target awareness and motivation, whereas training targets knowledge and skills. Both appear to be important in improving colonoscopy quality.]

Quality Gap: Colonoscopy quality is linked to patient outcomes. The national VA healthcare system has lacked a comprehensive program to measure and report colonoscopy quality and help providers improve adenoma detection.[17] Accurate measurement of quality metrics is challenging because validated quality metrics are not available in structured VA data from VA CDW. The challenge is multifactorial. Colonoscopy procedure documentation resides in text notes in Vista/CPRS or endoscopic reporting software. The commonly used VA endoscopy note writer software programs (i.e. Endopro®, Provation®, etc) do not facilitate tracking of pathology data and quality measurement. **[These same issues will persist with the move to Cerner Millenium® over the next decade unless an infrastructure is in place to collate 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.

Proposed Solution and Rationale for Study:
A primary justification for this study is that VA-EQuIP has a high probability of improving a quality metric outcome (ADR) directly associated with CRC death, one of the most common cancers in Veterans. A randomized controlled trial of this magnitude is an opportunity to show, for the first time, that even small improvements in clinical performance from audit and feedback, with support to improve skills and quality, can save Veterans lives, since even a 1% increase in ADR translates to a 3% reduction in interval cancer death. Our proposal for a 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 of VA-EQuIP implementation will identify factors associated with effective implementation and colonoscopy quality improvement at VA sites. Our prior and planned work builds toward our long-term goal to reduce mortality in Veterans by increasing early detection of CRC and inform national quality improvement initiatives such as remediation training for continual low performing endoscopists.

The novel foundation of VA-EQuIP is a natural language processing (NLP) algorithm and informatics reporting infrastructure that we developed to measure and report the quality of colonoscopies performed by individual VA colonoscopists.[33] [Our program is similar to the VA Surgical Quality Improvement Program (VASQIP) that houses a national quality assurance database for reporting, comparison, and quality improvement.[34,35] However, VASQIP requires significant manual labor with a dedicated nurse manager at every VA site to collect and enter data into the registry. Recreating the same model for gastroenterology procedures would be costly (if not impossible) and take years. 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.]

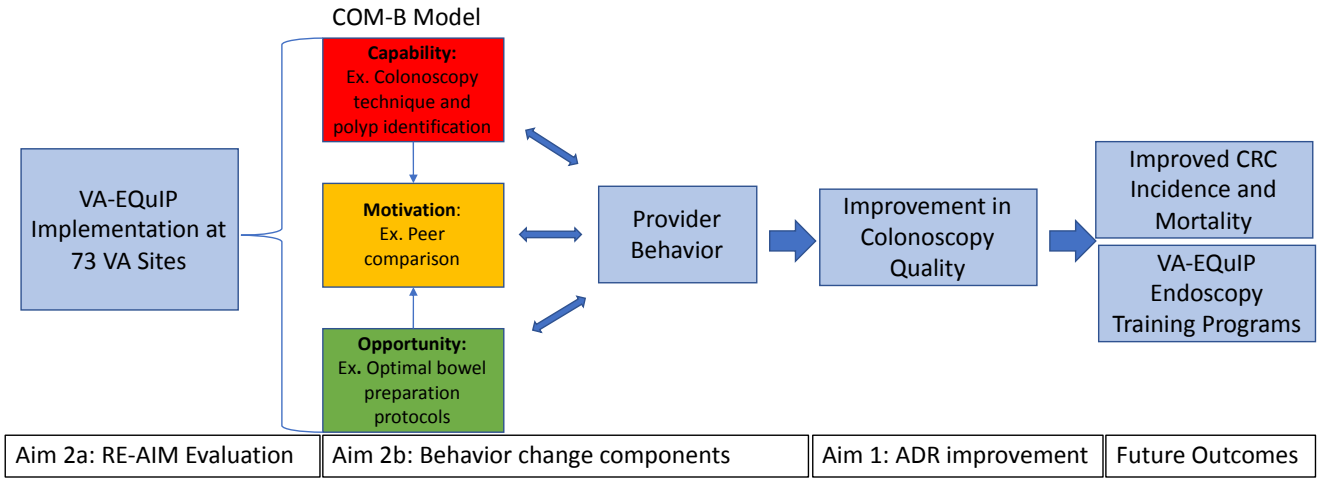
[With continued expansion to community care via the VA Maintaining Internal Systems and Strengthening Integrated Outside Networks (MISSION) Act, VA will be providing and paying for colonoscopy services by VA and non-VA providers. It is crucial that VA is able to collect, compare, and impact quality metrics from VA and community care that are linked by evidence to patient outcomes so that these measures of quality will be used to justify such crucial “make versus buy” determinations.] It is our hope that our study will not only improve colonoscopy quality but also provide a framework for assessing the quality of other specialty care procedures such as bone marrow biopsy or bronchoscopy.

[As prior evidence in the GI literature also supports education and training to improve ADR,[31,32] and taking into account the large number and heterogeneity of VA endoscopy sites and providers, VA-EQuIP’s audit and feedback implementation strategy will be supported by virtual collaborative learning sessions moderated by clinical and quality experts. This will promote active discussion, troubleshooting, and evidence based practice to improve quality. The approach avoids a “one size fits all” strategy such as passive webinars or costly hands-on training programs. The virtual collaborative learning sessions will provide a non-punitive, flexible forum to guide colonoscopy quality improvement in combination with our novel colonoscopy quality reporting.]

[Learning collaboratives were originally developed by the Institute of Medicine (“Breakthrough Series”) to support peer-to-peer and peer-to-expert learning to improve healthcare.[36] Specifically, quality improvement learning collaboratives allow sharing of expertise across practice sites to hasten the diffusion of evidence

based practices.[37] Our VA EQuIP collaborative will emphasize shared learning across multiple units led by nationally recognized experts in colonoscopy quality (Drs. Kahi, Kaltenbach, Gupta, Saini). Learning sessions will include didactic sessions, discussion, and skill building activities followed by periods where individual teams work on quality improvement projects.[37] 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.[38–40] Recently, "virtual" collaboratives have been described and used in many patient safety projects (fall prevention, pressure ulcer prevention, catheter infections) in the VA.[41–45] **Virtual collaborative learning has never been evaluated as an approach to improve colonoscopy quality. Our proposal will build on prior work to test, in rigorous fashion, if virtual collaborative learning combined with audit and feedback can improve an outcome measure (ADR) associated with CRC mortality. The results will expand knowledge on how best to operationalize virtual collaborative learning for clinical care in a large integrated, but diverse, healthcare system.]**

[Conceptual Model: Our conceptual model (**Figure 1**) incorporates the Capability, Opportunity, Motivation (COM-B) model of behavior change to guide and enhance our implementation strategy and evaluation.[1] We will also use the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework[46] with integrated Qualitative Evaluation for Systematic Evaluation (RE-AIM QuEST[47]). The evaluation is designed to target and capture constructs defined by REAIM-QuEST[47] as they apply to colonoscopy quality measurement, reporting, and improvement. Provider surveys and qualitative interviews will explore components of behaviors change mapped to COM-B domains.]



[Figure 1: Conceptual model of VA-EQuIP impact on behavior and colonoscopy quality]

Leveraging the informatics investment made by HSR&D as well as the national GI program office partnership in quality improvement initiatives, the roll out of VA-EQuIP is a rare opportunity to study, in real time, the deployment and impact of a large-scale learning health system initiative on provider behavior change and colonoscopy quality. This is where implementation science becomes essential, especially with respect to stakeholder engagement, adaptation and tailoring. Our team will evaluate VA-EQuIP implementation to better understand how and why such a program is effective or ineffective in different contexts, such as how to improve and enhance audit and feedback with collaborative learning.

SIGNIFICANCE

Our proposed evaluation is highly relevant to our operations partner and to national VA priorities. The VA gastroenterology community is leading national efforts to better define and deliver high-quality specialty care. The VA National Gastroenterology Program Office (Jason A. Dominitz, MD, MHS) is the sponsoring operational partner for VA-EQuIP and this initiative to implement and evaluate the impact on colonoscopy providers. Dr. Dominitz strongly supports this proposal and randomization method, and is committed to the evaluation. The proposed work clearly aligns with HSR&D and QUERI's goals of accelerating the timeline from evidence to implementation and impact, and utilizing hybrid study designs to derive generalizable knowledge from large-scale programmatic initiatives.

VA-EQuIP has the potential to quickly and dramatically improve variability in ADRs across VA and ultimately reduce the risk of CRC incidence and mortality. Our new data showing interval and fatal CRC

risk by ADR indicates significant variability in adenoma detection across VA with ADRs ranging from 13-79%.[23] VA-EQuIP will directly address the recent OIG recommendations for quality monitoring.[17] Moreover, quality improvements resulting from the proposed strategy have high potential to impact Veteran healthcare in a short time frame.

Secondary impacts of proposed research: The reported quality metrics could potentially contribute to ongoing professional practice evaluations (OPPE), which we will evaluate during the trial. Measurement of completeness (i.e. cecal intubation rate) and inspection quality (i.e. ADR) of colonoscopy is a recommended part of OPPE for VA endoscopists, and emphasizes to VA providers that VA values high quality procedures for Veterans. However, current data limitations and the time burden for manual data collection and processing are major impediments to providing and tracking accurate performance data to incorporate into OPPEs. **The reporting of quality measures will also improve access to care.** Colonoscopies with poor bowel preparation result in canceled procedures or repeat procedures at earlier intervals (e.g. 1 year). More repeat procedures leads to decreased access to endoscopy specialty care for new patients. If VA could identify and highlight high quality endoscopy centers and endoscopists, that data could be used to promote system wide improvement, protocol and documentation standardization, and adoption of best practices to ensure access to high quality colonoscopy for all Veterans.

Audience for research results and how they will use the products: VA-EQuIP will have multiple primary and downstream users and stakeholders including:

- The VA National Gastroenterology Program Office as noted above, the sponsor of VA-EQuIP.
- The Office of Specialty Care and Primary Care to ensure Veterans are referred to high quality providers.
- The Office of Veterans Access to Care (OVAC) and National GI Access initiative to ensure that access initiatives are aligned on providing Veterans with high quality procedures.

The proposal directly addresses VA priorities by system modernization and efficient use of resources.

1. **"**Modernize our systems": National standardized colonoscopy quality reporting will improve infrastructure and streamline services by providing accurate, prospective quality monitoring that currently does not exist.
2. Focus resources more efficiently: The manual chart review used by most sites to capture quality metrics results in significant unaccounted work burden across the VA healthcare system. VA EQuIP will make this inefficient practice obsolete.

Significance to Veterans and Veteran engagement: This proposal engages Veterans across the project spectrum. We solicited feedback from 12 Veterans on the Geriatrics Research Education and Clinical Center (GRECC) Veteran engagement panel, including a mix of persons who receive and do not receive VA care. Colonoscopy and colon cancer screening was very familiar to all Veterans present. A letter from the GRECC Veteran engagement panel (**Appendix 3**) highlights the support and excitement for the proposal. In particular, there was interest in expanding VA-EQuIP to allow patients to view provider colonoscopy quality data. At this time, VA-EQuIP is not "patient facing" but this Veteran feedback will inform future work in this area. At study conclusion we will reach out to the GRECC engagement panel and other Veteran groups to help disseminate research findings. In this way we can assure the effective reach of our research and also gain insights about the importance of our findings from Veterans themselves.

Completed work to inform implementation and evaluation: Multiple pre-implementation strategies[48] guided the planned implementation of VA-EQuIP by the GI program office. These strategies include our identification of the major barriers to colonoscopy quality reporting to help ensure effective implementation with audit and feedback as outlined below:

Needs Assessment: We published survey results showing marked variability in quality measurement and reporting practices at 93 VA facilities in 44 states.[18] **Importantly, ADR, the quality metric most strongly associated with risk of CRC mortality, is not being measured in over a third (38%) of sites.** A majority (87.8%) of respondents were interested in a centralized automatic colonoscopy quality reporting system to avoid performing manual calculations.

Barriers & Facilitators: We conducted 21 qualitative interviews of gastrointestinal (GI) section chiefs to understand barriers and facilitators in measuring and reporting colonoscopy quality**.** We identified six

predominant barriers to measuring ADR: 1) Time-consuming manual data retrieval; 2) Lack of dedicated time and resources; 3) Lack of standardized measurement; 4) Reliance on proxy measures; 5) Poor data resolution; 6) Need for automation for quality reporting. Our results are under review for publication (**Appendix 4**).

Develop and organize a quality monitoring system - Data and Infrastructure:
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 have optimized a machine learning document classifier that uses the support vector machine algorithm to delineate a note as either a colonoscopy procedure note, or not. Our process allows us to rapidly and prospectively confirm colonoscopy notes for NLP processing and colonoscopy quality reporting. Our NLP pipelines to extract colonoscopy procedure and pathology data have excellent performance (**Table 1**). The VA-EQuIP dashboard visually presents biannual and cumulative colonoscopy quality metrics to VA sites and providers, and provides comparison metric data at the local site and national level (**Figure 2**). All reports are created using Microsoft™ SQL Server Reporting Services 2013 and are hosted on a Microsoft SharePoint 2013 internal website. Access is controlled by both SharePoint memberships and local data access permissions.

Table 1: 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%

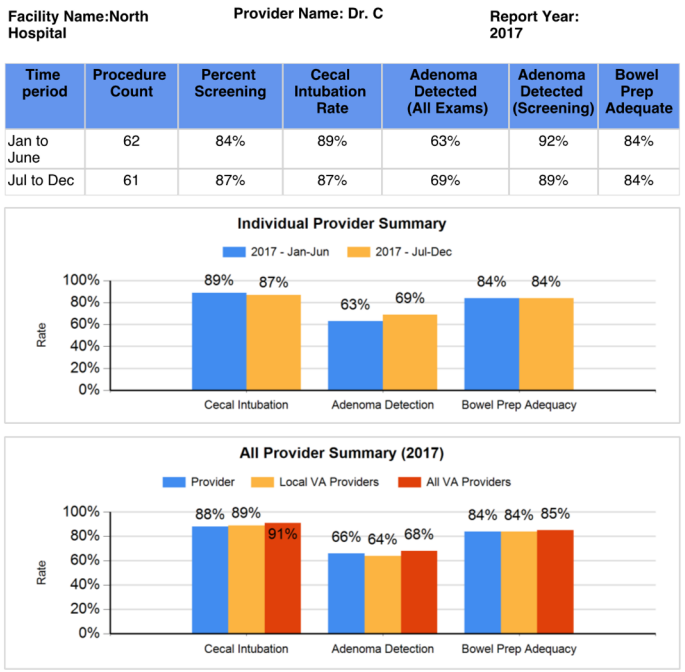


Figure 2: VA-EQuIP Quality Dashboard

Usability and [Pilot] Testing: In cooperation with our operations sponsor and partner, we obtained feedback and usability information on VA-EQuIP from the GI Field Advisory Committee. The committee members were provided access to their sites and individual data on colonoscopy quality. Respondents (N=5) provided feedback for internal consistency and were surveyed via the System Usability Scale (SUS).[49] The average SUS score for VA-EQuIP for 5 respondents was 87.0 (SD 9.9, range 77.5-100). A score >80.3 has been found to be in the top 10% of usability performance based on an evaluation of 500 prior evaluations using this scale.[50]

RESEARCH OVERLAP

We conducted a thorough search of multiple clinical trial databases and VA funded HSR&D projects using keyword searches **(Appendix 2).** The search resulted in the following: colonoscopy quality metrics (N=1 study), report card (N=30 studies), adenoma detection rate (ADR) (N=33 studies), and colonoscopy (N=641 studies). One prior VA study (RRP 12-184, PI Saini, July 2012-Sept 2013) evaluated electronic surveillance measures of overuse of CRC screening. A 10-year-old study (SHP 08-201, PI Lieberman, April 2008-Sept 2008) determined if colonoscopy quality indicators were being measured at 9 VA centers. Our searches appropriately identified the Measurement Science QUERI and IIR 14-092 (PI: Gupta, Optimizing Colorectal Cancer and Polyp Surveillance after Colorectal Polypectomy) that played a crucial role in infrastructure development for this current proposal. Detailed review of all studies revealed no overlap for implementation of a colonoscopy quality report card across the VA healthcare system with potential to impact >100,000 Veterans / year by improving CRC outcomes and reducing mortality.

RESEARCH DESIGN AND METHODS

Aim 1: To determine if VA-EQuIP implementation improves provider ADRs

[Overview of VA-EQuIP implementation strategy, study design, and timeline per randomized cluster
Our prospective, multi-center, stepped wedge cluster randomized trial design facilitates a graduated rollout of VA-EQuIP across participating sites, which is more feasible than a parallel randomized trial where an intervention is implemented across all sites simultaneously.[51] We will stratify randomization of sites to our six enrollment waves by site size (low/high annual patient volume) and site annual ADRs in 2018 (low/high). Low volume sites and low ADRs sites are likely to benefit the most from VA-EQuIP. If there were differences in site volume, or ADRs over time (ie, secular trends) and systematic differences in volume or ADR rates across enrollment waves, then these effects would be confounded in our analysis.[10] Thus, stratifying by them helps ensure that we can properly adjust for secular trends in our primary analysis. The national GI program office will implement VA-EQuIP with two primary implementation strategies of audit and feedback and collaborative learning. Other implementation strategies include the identification of a local quality steward who is the responsible site delegate for quality and centralized technical assistance throughout the study (**Table 2**). A visual timeline of VA-EQuIP implementation and evaluation per randomized cluster is shown in **Figure 3**.**]**

Table 2: VA EQuIP Implementation strategy

Implementation Strategy	**Operationalization per cluster (12-13 sites)**
Identify and prepare local quality stewards. Quality 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 local quality steward responsible for colonoscopy quality assurance and site level permissions for quality data - Local quality 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 *Local stewards: All site provider data access* o *Providers: Only individual data access* 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 quality 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

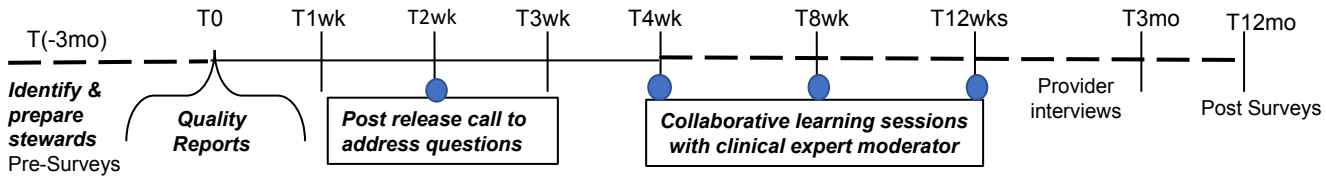


Figure 3: VA-EQuIP with strategies and evaluation timeline per randomized cluster

[Analysis Overview: Our goal is to compare the efficacy of VA-EQuIP versus usual care for the primary outcomes of overall (all indications) 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 rate of ADR, bowel preparation quality and cecal intubation relative to standard of care. We also hypothesize the effect will be particularly pronounced among providers who have low ADR, bowel preparation quality and cecal intubation rates 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, which is commonly used to study intervention effects when an intervention occurs at different times within different providers. 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.]

Eligibility / Inclusion criteria: We have identified 73 VA sites spanning VA (**Figure 4**) eligible for the study that have endoscopists with colonoscopy procedure and pathology notes in our database (SCS_Endoqual). In 2017, these 73 sites had 657 providers performing 155,926 colonoscopy procedures. The point of contact (i.e. GI section chief) at each site will be contacted by email about planned VA-EQuIP implementation and given specific instructions for accessing site data.

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

Consent: The national GI program office will implement VA-EQuIP as an operational initiative. We will obtain provider verbal consent by telephone, if deemed necessary by the IRB, for the evaluation.

Figure 4: VA-EQuIP sites

Colonoscopy quality report card & reporting infrastructure: The VA-EQuIP quality reporting infrastructure includes 2 separate CDW databases with an extract, transform, loading (ETL) framework and secure SharePoint Site to visually present quarterly colonoscopy quality metrics (**Figure 2)**. VA-EQuIP has been pilot tested and undergone iterative usability testing with the Gastroenterology Field Advisory Committee.

Data measures: CDW structured data include patient demographics and endoscopist name and specialty (e.g. gastroenterology, surgery). Variables derived from natural language processing of colonoscopy procedure and pathology notes include extent of exam, screening indication, and adenoma detection.

Outcome measures: 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 defined as the percentage of patients undergoing colonoscopy for any indication with one or more adenomas detected. Screening ADR is defined as the percentage of patients ≥50 years undergoing screening colonoscopy who have one or more adenomas detected. We will assess secondary outcomes measures, including cecal intubation rate and bowel preparation quality. Cecal intubation is defined as passage of the colonoscopy tip to a point proximal to the ileocecal valve so that 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.

Primary Analysis -- Stepped Wedge at the Provider Level: We will utilize patient data aggregated to the provider level at monthly intervals to examine how the ADR changes pre- and post-intervention. A provider's monthly ADR is the ratio of the total number of the provider's positive ADR screening results to the total number of patients that the provider screened during the month. Prospective data collection will begin Jan 1st, 2020 and conclude Jun 30, 2022, a 30-month data collection period. Baseline data from 2019 will be included in our analysis to augment the pre-intervention time frame, to enable our intervention to start immediately. ADR

rates for 2018 will be used to 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). For each wave, we will discard the data from the 3-month period where the intervention is being implemented. We will augment baseline data capture by 12 months and we will extend data capture following our final wave by another 12 months to improve estimation of pre/post intervention effects within individual providers who are members of the first and final waves. Under this design, both the pre- and post-intervention data available will range from 12-27 months depending on the wave, with an average of 19.5 months of data.

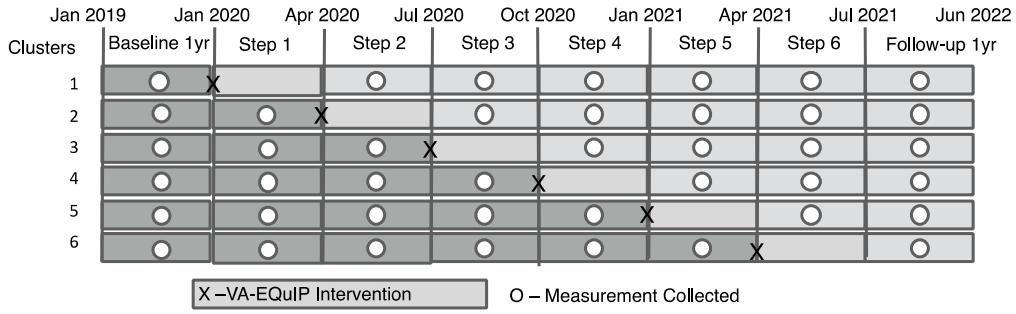


Figure 5: Enrollment Strategy for Stepped Wedge Trial Design

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. We will use a Poisson outcome model with a log link to characterize each provider's monthly rate, where the number of positive screens will be the outcome and the log of the number of total screens will be included as an offset term.[53,54] In the case of over-dispersion, 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 (Intervention). We will additionally control for time since the study began in quarters (*Time*, coded as a categorical variable) to account for secular trends; as well as site level characteristics obtained from administrative and the quantitative survey data (Aim 1) (urban/rural, cumulative annual procedure volume in 2018 and current volume, site annual ADRs in 2018, use of split dose bowel preparation protocols (SDBP) in 2018, and whether or not ADR was measured in 2018), provider level characteristics (baseline years in practice, primary practice type {academic, hospital employed, private practice}, and baseline annual procedure volume), and baseline patient characteristics aggregated to the provider level (age, sex, race/ethnicity). Since study time will be coded in quarters, it will 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 the ratio of 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 how the ADR rate changes over time, and to check for the presence of a lagged intervention effect. We expect to see about a one month lag in the intervention effect and plan to drop the outcome data for this period to avoid biased estimates.[55] 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.[56] If these 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 under the assumption that outcomes are independent between sites, and use robust standard errors for statistical inference.[57] The GEE approach can also be used as a fallback if the nested random effect structure of the generalized linear mixed model turns out to be numerically intractable. [As a secondary question we will assess the impact of compliance with the intervention on ADR rate among the providers from Aim 2 who responded to our survey. Compliance will be measured in Aim 1 by dashboard access and participation in one or more 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. Based on 2017 data there were 73 sites with 657 providers performing 155,926 cases per year in our sampling frame. We estimate a 5% loss of site data due to data quality issues. We also estimate a provider attrition rate of 8% from four years of data provided by the national GI program office where attrition rates ranged from 6.9-12%.[15] After accounting for data loss and provider attrition, we expect 69 sites with 571 providers seeing 135,517 patients/year, corresponding to 95 providers/wave and 59 patients/provider/quarter. Note that 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 is ADR, which we will conservatively estimate as having a range of 13-79% based on prior research[7,8] 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.[58] Based on an ADR of 27.5% in the pre-intervention period, we would 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.[7] 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.[59,60]

Secondary Analysis -- Interrupted Time-Series (ITS) at the Provider Level: 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. As a secondary analysis, we will estimate the average intervention effect across providers using a two-stage approach. In the first stage, we will implement ITS models within each provider, estimating the intervention effect for each provider. In the second stage, we will use a meta-regression approach to combine results across providers to estimate the mean and the standard deviation (on the log scale) of the intervention effect across providers.[61,62] This analysis is conventional for pre/post analysis of interventions implemented at different time points for different units,[62,63] and has the advantage of evaluating both the overall intervention effect and the heterogeneity in the treatment effect across sites and providers. However, it focuses on pre/post comparisons within providers and does not allow for comparisons between treatment and control at the same time points as is done in our stepped wedge analysis.

The first stage ITS analysis will consist of segmented regression with an autoregressive error model to account for the correlation between successive monthly ADRs. As described in the primary analysis, we will plot monthly ADRs for a random sample of endoscopists to guide our ITS modelling strategy. We will also use these plots to assess whether to include seasonal effects, which would be included as a categorical quarter variable in the model. 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: $\text{Log}(PositiveScreens_t) = \beta_0 + \log(TotalScreens_t) + \beta_1 Intervention_t + \beta_2 Time_t + \beta_2 Time\text{-}After\text{-}Intervention_t + \beta_8 Age + \beta_9 Sex + \beta_{10} Race + \beta_{11} Season_t$, where t indicates time coded in months.[55] In this ITS analysis, time is coded monthly and treated as a continuous variable instead of coding it as a categorical variable as in the stepped wedge analysis. 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.[56] 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.[61,62] This model will enable us to estimate the average intervention effect across providers accounting for heterogeneity at both the site and provider level by including random effects for both. We include a random effect for provider because although we have included the majority of the eligible providers in the VA system, we conservatively consider the providers included in our analysis to be a random sample from the full possible set of providers.[64] Konstantopoulos et al. describe a similar multilevel meta-regression analysis where studies are nested within school districts – here we have providers nested within sites.[65] Our meta-regression model will adjust for the variables used in the randomization stratification procedure (*Volume2018, ADR2018,* and *ADRMeasured*), provider-level variables (baseline years in practice and baseline annual procedure volume), and site level characteristics including [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 after accounting for 5% data loss at the site level and 8% provider attrition. On average a provider has 19.5 months of data in both the pre-intervention and post-intervention periods. Thus we estimate 386 patients/provider or 220,215 patients seen in each period. Again, the number of patients is included in the power calculation because the number of patients seen by providers greatly affects the power of our design. To estimate power, we simulated ITS data with varying intervention effect sizes adapting code provided by Rozario et al.[66] 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 – Stepped Wedge and ITS analyses at the Patient Level: One potential limitation of the analyses described above is that we are aggregating patient-level data to the provider month-level. The analysis of aggregated data 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.[67,68] This is due to insufficient adjustment for confounders when the data is analyzed at the aggregated level. The ITS analysis is particularly vulnerable to Simpson's paradox as there is no comparison of intervention and controls at the same point in time, but rather this analysis focuses on pre/post comparisons within providers. Thus, there is potential for confounding due to case-mix differences in patients seen at different points in time. To overcome this paradox, we will re-examine both our primary stepped wedge analysis and our secondary ITS analysis using data at the patient observation level. The modeling will be very similar, where we will use Poisson outcome models, but there will be no offset term at the patient level. The interpretation of the results will be the same as before. For the stepped wedge analysis, the exponentiated coefficient of intervention will represent the ratio of the screening rate after implementing the intervention versus before, for a given provider. For the ITS analysis, the first stage modeling will yield the ratio of each providers screening rate after implementing the intervention versus before. The second stage will yield a pooled estimate of the intervention RR across providers.

Aim 2: Evaluation of VA-EQuIP implementation

[Evaluation Overview: We will use a mixed-methods approach to perform the evaluation. This multi-faceted evaluation consists of three components: monitoring of VA-EQuIP usage statistics, pre/post surveys of quality stewards and providers, and qualitative interviews with quality stewards and individual providers. The approach will capture key RE-AIM (1a) and COM-B (1b) domains with qualitative assessments to explore barriers and facilitators to implementation and components of provider behavior change.**]**

Evaluation Measures and Outcomes:

Aim 2a: The evaluation measures will include reach, adoption, implementation, and maintenance domains (**Appendix 5**). Reach will be measured with site practice characteristics items from the baseline quality steward survey. We will also evaluate claims data from CDW fee basis, VA Choice and MISSION Acts, and contracting records to determine the proportion of Veterans at each facility who have colonoscopies performed by VA versus non-VA (community care) providers (another reach measure). Our comprehensive data infrastructure allows us to electronically monitor VA-EQuIP quality data views by site and provider (our primary adoption measure). These data will be used as quantitative, continuous measures of adoption at both the site and individual provider level. We will also collect participant data from those sites and providers participating in collaborative learning sessions. We will also include additional adoption measures in the quality steward

surveys by asking about use of VA-EQuIP for quality measurement and reporting at each site. As for implementation measures, the centralized VA-EQuIP infrastructure will also allow us to capture fidelity by assessing timing of delivery of the intervention (fidelity). The quality steward surveys were designed to capture outcome measures for sites (e.g. change in bowel preparation protocols) before and after VA-EQuIP implementation (**Appendices 6-7**). Qualitative interviews with quality stewards will explore site level barriers and facilitators to VA-EQuIP implementation and start after the last collaborative learning session for each randomized cluster.

[Aim 2b: The provider surveys and qualitative interviewers were designed to capture components of behavior change mapped to domains of the COM-B model (e.g. capability: knowledge of colonoscopy techniques to enhance adenoma detection) before and after VA-EQuIP implementation (**Appendices 6-7**). Provider qualitative interviews will start after the last collaborative learning session for each randomized cluster.**]**

Quantitative Survey Design and Qualitative Interview Guide: [Our quantitative surveys for quality stewards and individual providers include measures we have defined for each domain. Surveys will be delivered pre- and post (12 months) VA-EQuIP implementation; interviews will be conducted after each randomized cluster through trial completion.] A full description of all data sources, measures, outcomes, quantitative surveys, and qualitative interview guide are available in **Appendices 5-8.** We have developed draft survey questionnaires for quality stewards [and providers and a provider interview script to capture the intended constructs of our conceptual framework.] Because our measures are specific to the VA-EQuIP intervention, it was necessary to develop specific questionnaire items with our co-investigator and survey design expert Dr. Morgan Millar. Development has been an iterative process, wherein all necessary constructs and domains were identified and appropriate questions for capturing these domains were constructed (see draft, **Appendix 6-7**). The quality steward surveys will assess site characteristics (RE-AIM) and colonoscopy protocols before and after VA-EQuIP implementation for use in assessing site level change after implementation. Practice changes include whether or not sites changed clinical protocols or practice, including quality measurement practices, quality reporting practices, bowel preparation protocols, endoscopy technique (e.g. withdrawal time), use of assist devices marketed to improve adenoma detection (e.g. cap), or seeking additional training or proctoring. The provider survey was modified from a previously developed, validated COM-B-based survey[69] 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).[69] Qualitative interviews with providers will explore components of behavior change related to VA-EQuIP implementation (**Appendix 8**).

Quantitative Survey Administration: The quantitative surveys will be delivered to the identified local quality stewards (N=73) and providers (N=657) for each enrolled site via the Research Electronic Data Capture (REDCap)[70] platform at baseline (prior to VA-EQuIP) and 12 months following completion of the VA-EQuIP intervention. We will incorporate established methods of maximizing web survey responses, including multiple, carefully-timed, integrated email contacts.[71]

Qualitative Interviews and Transcription: Trained interviewers, supervised by Dr. Susan Zickmund, will conduct and audio record up to 60-minute telephone interviews with and individual endoscopists at participating facilities. All interviews will be transcribed verbatim by the VA Health Services Research and Development (HSR&D) Centralized Transcription Services Program (CTSP), directed by Dr. Zickmund and located in Salt Lake City. Purposive sampling[72] will create the sample needed for interviewing endoscopists at the facilities. Based on principles of thematic saturation (the concept that no new themes emerge from subsequent interviews), we will estimate a minimum sample size of at least 50 persons **(25 quality stewards and 25 individual providers)** across the sites enrolled in our study, 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 has been chosen to ensure saturation.[73] We will continue with interviews until thematic saturation has been reached.[73]

Analysis

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

Survey analysis (Baseline and 12 month outcomes): [Survey data analysis will include descriptive statistics of quality steward and individual provider survey responses at two timepoints: baseline and 12 months after VA-EQuIP implementation. These will include response distributions/percentages or means and standard deviations for all baseline site characteristics and all implementation and behavior change measures. We will examine what percent of providers viewed their quality data and if site-level activities that could impact provider behavior ("opportunities") had been initiated after the intervention. Using the quality steward survey, we will assess the proportion of sites exhibiting changes in protocols and be able to assess sustainability of new practices invoked earlier in the intervention. We will summarize responses for the provider survey using the scoring approach from the COM-B survey that we adapted for our provider survey, and will compare responses from the baseline and 12-month surveys. We will also compare COM-B measures between providers, stratified by ADR categories (e.g. quintiles) pre- and post-VA-EQuIP and by involvement in collaborative learning as a categorical variable.]

Qualitative analysis: Qualitative interviews will be analyzed using the verbatim transcripts for codebook construction using the qualitative "editing" method by Crabtree and Miller.[74] Representative quotations will be captured verbatim using Atlas.ti (Scientific Software, Berlin Germany) [and mapped to COM-B behavior change domains.] Working with Dr. Zickmund, two extensively trained qualitative analysts will meet and process any differences until they agree. The codes determined through this agreement process will be recorded in a master file and become the basis for the final analysis. This process of coding independently (the basis for the inter-coder reliability scores) and then discussing each case will ensure narrative coherence in the coding process. Inter-coder reliability kappa scores of below 0.61 will trigger a training meeting. This process has consistently achieved an inter-coder reliability kappa score of ≥ 0.70.[75]

NEXT STEPS

Standardized, transparent, automated quality reporting is the first step to ensure and improve colonoscopy quality. [Due to the sensitive nature of provider quality metrics, heterogeneity of VA sites, and desire to maintain local autonomy, we postulate that the collaborative learning sessions will help drive peer-to-peer and peer-to-expert learning to impact quality improvement in a non-punitive forum.] The interactive problem solving and support in our collaborative learning represents a limited form of external facilitation. **A critical next step will be to provide access to more intensive interventions, such as site visits and hands-on training, for persistently low performing providers based on findings from our evaluation.** Research has demonstrated that colonoscopy quality can be improved through a short course to train colonoscopists how to perform a higher quality exam.[31] The national GI program office is still planning a joint training program with VA's Employee Education System (EES) for colonoscopists in order to improve the quality of colonoscopy in VA and ultimately, Veterans' health outcomes. The training will take advantage of lessons learned from the Kaminski study[31] and would be co-directed by Drs. Tonya Kaltenbach (Co-PI) and Charles Kahi (Co-I) in collaboration with leadership of the SimLEARN Center (Orlando, Florida). Highly qualified faculty will be recruited from across VA to lead a 1.5-day course comprised of didactic lectures and hands-on skills enhancement for approximately 50 VA colonoscopists. Future training programs of this type are likely to be needed given the large number of providers performing colonoscopy in VA. Adding hands-on-training to VA-EQuIP will be the critical next step for improving refractory colonoscopy quality deficiencies identified after VA-EQuIP implementation. An additional next step is to report over and underuse of colonoscopy to improve access to GI specialty care. The National GI Access initiative (Dr. Susan Kirsh) is developing measures of access and overuse of colonoscopy. In collaboration with Dr. Kirsh's office, VA-EQuIP can ensure that access initiatives are focused on providing Veterans with high quality procedures. If validated overuse measures become available in the VA-CDW we would be able to incorporate them into VA-EQuIP.

DISSEMINATION PLAN

Our strong partnership with the Measurement Science QUERI and national GI program office will provide ongoing dissemination throughout the project via print and electronic publications. We will post news about progress of the project on both the Measurement Science QUERI Web Site and National GI Sharepoint Site. We will alert VA stakeholders to new postings via electronic internal news releases. At the conclusion of the project, Drs. Gawron and Kaltenbach will collaborate with Dr. Dominitz to produce a VA HSR&D CyberSeminar to notify critical VA stakeholders of the findings. Our leadership team and members of the GI Field Advisory Committee will participate in the seminar development. Specifically, they will help formulate the implications from the study and key members of our team will be available during the CyberSeminar to answer questions.

All slides and formal notes will be posted on the Gastroenterology Program SharePoint site, ensuring their continued availability for VA providers. We plan similar presentations for the American Gastroenterology Association and American College of Gastroenterology. The academic and research communities beyond VA will be reached through traditional means, including publication in high impact peer-reviewed journals specific to the field, such as *Gastroenterology* and the *American Journal of Gastroenterology*. Results will be presented at national conferences hosted by VA HSR&D and AcademyHealth. Policy makers and others (e.g. Office of Specialty Care) in a position to affect clinical care will be reached through a policy brief, which we will distribute in collaboration with the national GI program office to researchers, federal, state, and local elected leaders, and Veterans' organizations.

SAMPLE RECRUITMENT

Recruitment timeline / detailed monthly plan for sample recruitment: The implementation strategy timeline and monthly plan for the trial is shown in **Figures 3 & 5**. VA-EQuIP dashboards will be delivered in 6 waves of 12-13 sites (N=73). We will deliver our quality steward and provider surveys before and at 12 months after each cluster in our stepped wedge design to study completion. We will recruit endoscopists for qualitative interviews starting at 12 weeks (after VA-EQuIP intervention) and conduct 2-5 interviews per cluster until study completion.

Pitfalls / Alternative plans if sample recruitment does not meet expectations:

For Aim 1, we have accounted for provider attrition and problematic data in our power calculation, based on our own and historical data provided by the national GI program office. The centralized infrastructure and NLP pipelines of VA-EQuIP will allow creation of reports for all providers in our database to ensure that the NLP system performance remains high throughout the project, several internal validation studies are planned. Each validation will involve a manual annotation of a random set of notes selected at each of the participating sites to create a reference standard. The system performance metrics will be calculated by comparing NLP system output and the reference standard. Since the NLP system was extensively evaluated during development, validation will be limited to measuring precision of extraction. If precision of extraction of each of the variable stays above 90%, the NLP system will be applied on the colonoscopy reports without changes. If evaluation detects decreased precision, detailed error analysis will be performed and system adaptation might be required. If system adaptation is performed, another round of validation will be required. In this case, both precision and recall will be evaluated. Colonoscopies performed on Veterans by providers outside of VA could be assessed for quality metrics using the NLP system contingent on the future integration of the colonoscopy and pathology data into VA CDW.

For Aim 2, our pre-implementation work provides strong evidence that our team is capable of performing the evaluation and maximizing survey response and qualitative interview completion. Our published needs assessment survey[18] had a response rate of 69% facilitated by 3 electronic reminder emails. We anticipate an even higher response rate for this project as the prior work required updating of outdated contact information, which is now available to us. For the qualitative work to determine barriers to quality reporting we obtained 21 interviews in <3 months with only single email invitations to VA sites. If survey response rates are not meeting expectations with electronic reminders, we will contact the local stewards by telephone to assess technical or other issues with the survey. The GI program office will promote and encourage survey response and interview follow up. A limitation of our study is that we are not currently able to measure quality for community care colonoscopies, yet we do plan on assessing the proportion of community care colonoscopies performed for future efforts to incorporate community care colonoscopy quality into VA-EQuIP.

PROJECT MANAGEMENT PLAN

Drs. Gawron and Kaltenbach, who have worked closely together over the past five years with multiple site visits between their VAs, will be responsible for overseeing all aspects of the project, including achievement of the project aims, conducting the quantitative and qualitative analyses, timely provision of project deliverables, coordination of the project team and workgroups, incorporation of feedback from the Steering Committee, and facilitation of partnerships with the national GI program office and FAC, QUERI, and other VA constituents. The Steering Committee (Drs. Gawron, Kaltenbach, Dominitz, [Helfrich], [Samore], Whooley, and Saini) will meet virtually monthly to review all major decisions, monitor the project's progress in meeting recruitment and other goals, to review the data analysis and findings, and coordinate publications and other products. The project aims will be advanced by 3 workgroups. Inclusion of Drs. Gawron and Kaltenbach on each workgroup will

ensure communication across the project. Each workgroup will meet weekly to biweekly to advance relevant tasks. The full team will meet monthly to coordinate efforts across workgroups and synthesize preliminary results. Monthly team meetings will invite Dr. Dominitz (GI program office) or a GI FAC representative designated by Dr. Dominitz.

The Quality Reporting and Informatics Workgroup, chaired by Dr. Gawron, includes Drs. Kaltenbach, Patterson, Gupta, 3 data programmers, and VINCI support staff. This workgroup will be responsible for ongoing updates and maintenance of the VA-EQuIP database, NLP processing of colonoscopy procedure and pathology notes, and maintenance of the reporting infrastructure. This will also include granting appropriate permissions to each site and provider to view VA-EQuIP data.

The Implementation and [Collaborative Learning] Workgroup, chaired by Dr. Kaltenbach, includes Drs. Kahi, Gupta, Helfrich, Whooley, Saini, Kahi, and Dominitz. This workgroup will be responsible for ensuring a broad range of stakeholder engagement and leading the collaborative learning sessions outlined above. Drs. Whooley and Helfrich will ensure that appropriate implementation strategies are applied to both anticipated and unanticipated challenges.[76] For example, tailoring of the report card may be necessary for sites that have a small number of endoscopists due to concern about individual providers being identified by their peers. Such adjustments will be based on the key principles of balancing fidelity with adaptability of the intervention. Drs. Kaltenbach, Kahi will lead the collaborative learning sessions with assistance from our other co-investigators and experts (Saini, Gupta, Gawron).]

The Evaluation Workgroup, chaired by Dr. Helfrich, includes Drs. Gawron, Zickmund, Taber, Whooley, Presson, Millar, a data programmer, and a program manager. This workgroup will be responsible for reviewing and final refinement of the quantitative surveys and interview scripts, and guiding delivery of the survey and interviews. They will also review the interview transcripts and all new qualitative codes for codebook inclusion.

Research Team and Relevant Experience: As described above, we have a strong record of working with our operational partner to improve colonoscopy quality. Proposal investigators bring expertise in many areas that will benefit the successful completion of the proposed work. Drs. Gawron (Co-PI) and Kaltenbach (Co-PI) have led the initiative (Measurement Science QUERI, Project 1) to build the infrastructure for VA-EQuIP - the foundation of this proposal. Dr. Gupta (Co-I) is PI of a VA Merit Review focused on developing a prediction model for CRC risk after colonoscopy. Drs. Gupta, Kaltenbach, and Dominitz have led randomized trials. Dr. Kahi, a VA national leader on colonoscopy quality, brings strong content expertise to the project, as he has developed and implemented a single institution VA report card that is a model for this study.[2] Dr. Zickmund (Co-I) is Director of the HSR&D Centralized Transcription Services Program with expertise in qualitative methodology. Dr. Whooley is the PI of the Measurement Science QUERI and experienced in implementation science and evaluation methods. Dr. Sameer Saini (Co-I) has expertise in implementation science and is PI for another randomized program evaluation using similar methods (focused on proton pump inhibitor de-escalation). **[Dr. Christian Helfrich has extensive experience in leading implementation programs and with a focus on the adoption of new, evidence-based practices. Dr. Matthew Samore will now serve on the project steering committee and brings a breadth of HSR&D experience including as the PI of both the Salt Lake VA Informatics Decision Enhancement and Surveillance Center and the VA Consortium for Healthcare Informatics Research.]** Our methodology Co-Is include statistical experts in design of cluster randomized trials (Dr. Presson and Dr. Greene), survey design (Dr. Millar), and natural language processing (Dr. Patterson). An overall study timeline with milestones is shown in **Figure 5**.

Figure 5: Study timeline

		Year 1				Year 2				Year 3			
	Q4	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
VA EQuIP Implementation		•	•	•	•	•	•						
Aim 1: ADR change													
Outcome measurement				•	•	•	•	•	•	•	•	•	
Aim 2: Evaluation													
Baseline surveys	•	•	•	•	•	•							
12 month surveys						•	•	•	•	•	•		
Qualitative interviews			•	•	•	•	•	•	•				
Final analyses												•	•