COVER PAGE

Study Title: Prospective RCT Of Water Exchange (WE) vs. WE Plus Cap-Assisted Colonoscopy

Unique Protocol ID: GAST-015-16S

ClinicalTrials.gov ID: NCT03160859

Document Dates: Protocol & Statistic Analysis (as approved): 8/14/2017

Record Verification: October 2023

Overall Status: Completed

Study Start: June 8, 2018 [Actual]

Primary Completion: October 28, 2022 [Actual]

Study Completion: October 28, 2022 [Actual]

Sponsor: VA Office of Research and Development

Responsible Party: Sponsor

Collaborators: (1) VA Northern California Health Care System; (2) VA Palo Alto Health Care System

General analytic plan (data management plan, sample size calculations, statistical analysis):

Data management plan. During the planning and start-up period, all elements of an accurate, reliable, and efficient data management plan will be developed by the Data Curation Center (DCC). The database programmer, in close collaboration with the cross-functional DCC team and with input from the investigators, will build the study-specific data collection application and database. OpenClinica® [www.openclinica.org] is a clinical trial software platform for electronic data capture, clinical data management, and clinical monitoring that is a 21 CFR Part 11 compliant system, using Oracle® as the relational database back-end. We will build the infrastructure necessary to input research data into the Oracle® databases using a web-based data entry system that incorporates front-end edit checks (e.g., data range and logical checks) to ensure that data are entered completely and correctly. Additionally, the DCC data manager and site study coordinators will develop a plan to identify and resolve back-end data errors. The enhanced discrepancy management function in OpenClinica® will expedite query resolution by allowing electronic communication of discrepancy reports between the UCLA DCC and site coordinators. The DCC programmer will develop the web-based randomization application. The DCC will develop consistency checks; both within patients across study visits and between sites. The clinical monitoring and data management procedures will be consistent with the International Conference on Harmonization (ICH E6) standards for Good Clinical Practice. Quality assurance will be accomplished in part by monthly data monitoring reports that will be discussed during meetings of the Steering Committee (PD/PI, site PIs and statistics consultant).

Sample size determination and description of subject population.

Updated male Veteran sample size. Pilot data in scheduled unsedated male Veterans were updated. The real time maximum insertion pain (RTMIP) [mean (SD)] was 2.7 (2.2) with WE (n=87) and 1.5 (2.1) with WECAC (n=38). The reduction (44%) 2.7-1.5=1.2 is 1.2/2.2=0.55 standard deviation (SD) units. The sample size/group needed to detect a difference at a level of significance of 0.0492 (adjusted for Obrien-Fleming interim stopping rules, one interim analysis at 50% completion) with 90% power is 72 or 144 total using a two-sided two sample t-test. Assuming 90% of subjects will be evaluable, we plan to recruit 160 male Veterans. Updated pilot data showed the PI performed 30 scheduled unsedated colonoscopies (male) in 1 year when ~2200 colonoscopies were offered at VAGLAHS (1.36%). At participating sites, a total of ~9000 colonoscopies (VAGLAHS 2200, VANCHCS 3300, VAPAHCS 3500) will be performed each year. For purpose of conservative estimation, 1/2 of 1.36% (0.68%) of all colonoscopies will be presumed to be in scheduled unsedated male Veterans when the option is made available up front. In 51 month (4.25 year) enrollment period, a total of 38250 colonoscopies will be performed. 260 (0.68% of 38250) male Veterans will receive scheduled unsedated colonoscopy, making target enrollment of 160 achievable (only 160/260 = 62% enrollment rate needed). In support of this possibility, in two prior RCT in scheduled unsedated male Veterans at VAGLAHS and VANCHCS comparing WE vs. air colonoscopy, high enrollment rates of 97.7% (84/86) and 95.2% (100/105), respectively, were recorded. Sample size of 160 will be sufficient to show significant differences in some of the previously listed secondary outcomes related to quality colonoscopy in male Veterans (e.g. proportion with no pain, insertion time, proximal colon ADR). The remaining previously listed secondary outcomes in male Veterans will be dropped (e.g. recalled pain, successful cecal intubation, overall and right colon ADR, willing to repeat).

Secondary outcome power estimates based pilot data in WE (n=87) and WECAC (n=38) in male Veterans

Secondary Outcomes	Pilot (WE vs. WECAC)	Difference (SDU or %)	Sample size per group needed*
Recalled pain, score (SD)	1.5(2) vs. 0.5(2)	1 (0.5 SDU)	84 (Will be dropped)
Successful cecal Intubation	98% vs. 100%	2%	(Will be dropped)
Proportion with no pain	18% vs. 46%	28%	54
Insertion time, minutes (SD)	15(5) vs. 12(5)	-3 (0.6 SDU)	58
Overall ADR - entire colon	36% vs. 50%	14%	258
Proximal colon ADR	17% vs. 40%	23%	76
Right colon ADR	20% vs. 40%	20%	106
Willing to repeat, score (SD)	9.0(2.5) vs. 9.6(0.9)	0.6 (0.4 SDU)	(Will be dropped)

ADR, adenoma detection rate; SD, standard deviation; U, units. *90% power, 95% confidence or 5% α error level.

Updated female Veteran sample size. After receipt of the critique, the PI identified 9 scheduled unsedated female Veterans (5 WE, 4 WECAC). RTMIP were WE, 6.2 (3.5) and WECAC, 3.5 (3.7). The reduction (44%) was 2.7/3.7 or 0.7 SD unit. The female Veterans had significantly higher RTMIP than the male Veterans when examined by WE, but the percent reduction of RTMIP by WECAC was equivalent (44%). The sample size/group needed to detect a difference at a level of significance of 0.0492 (adjusted for Obrien-Fleming interim stopping rules, one interim analysis at 50% completion) with 90% power is 43 or 86 total using a two-sided two sample t-test. Assuming 90% of subjects will be evaluable, we plan to recruit 96 female Veterans. Sample size of 96 will be sufficient to show significance in differences in some of the previously listed secondary outcomes related to quality colonoscopy in female Veterans (e.g. proportion with no pain, insertion time, overall and segmental ADR). The remaining previously listed secondary outcomes in female Veterans will be dropped (e.g. recalled pain, successful cecal intubation, willing to repeat).

Secondary outcome power estimates based pilot data in WE (n=5) and WECAC (n=4) in female Veterans			
Secondary Outcomes	**Pilot (WE vs. WECAC)**	**Difference (SDU or %)**	**Sample size per group needed***
Recalled pain, score (SD)	3.1(3) vs. 1.6(3)	1.5 (0.5 SDU)	84 (Will be dropped)
Successful cecal Intubation	100% vs. 100%	0%	(Will be dropped)
Proportion with no pain	0% vs. 25%	25%	32
Insertion time, minutes (SD)	17(4.5) vs. 14(4.5)	-3 (0.7 SDU)	43
Overall ADR - entire colon	20% vs. 50%	30%	48
Proximal colon ADR	20% vs. 50%	30%	48
Right colon ADR	20% vs. 50%	30%	48
Willing to repeat, score (SD)	9.2(1.1) vs. 9.5(1.0)	0.3 (0.3 SDU)	233 (Will be dropped)
ADR, adenoma detection rate; SD, standard deviation; U, units. *90% power, 95% confidence or 5% α error level.			

In both male and female Veterans, record will be kept of all these variables, however, to demonstrate comparability of the study and control groups. When the male and female data are combined, the sample size may reveal differences between WE and WECAC that are worth pursuing, thus providing pilot data for future studies (e.g. overall and segmental adenoma detection rates in scheduled unsedated subjects).

Statistical Analysis. Analysis will be performed based on assigned groups using STATA software (StataCorp 2013. Stata Statistical Software. College Station, TX: StataCorp LP). Descriptive statistics (frequencies, measures of central tendency and dispersion) will be used to evaluate the distribution of each variable. Data as mean (SD) or median (IQR) will be assessed by 2-samples independent samples tests (e.g. Student's t), Fisher exact test; relative risk tests and correlation analysis will be used to evaluate hypotheses. $P<0.05$ will be considered significant. Statistical assumptions required will be checked and appropriate data transformations or non-parametric test (e.g. Wilcoxon-Mann-Whitney *U test*,) will be used if assumptions are not met. Linear regression will be used to determine significant factor(s) that impact on the primary outcome. The analysis will use intention-to-treat principles, i.e., all randomized patients will be included in the analysis even if their colonoscopies were discontinued early (e.g., pain, poor preparation). The pain score at discontinuation will be used for the analysis for patients with early discontinuation.

Hypotheses Testing. There will be one-interim analysis and a final analysis for this study. The interim analysis will be conducted after 50% of subjects have been evaluated. This interim analysis will be for the primary analysis alone. The significance level for the interim analysis will be 0.0054 based on the standard O'Brien-Flemming rule. The study will be stopped if there is a significant result for the interim analysis. The final analysis will use a 0.0492 level of significance.

Primary hypothesis. Real-time maximum insertion pain (RTMIP) is significantly lower in the patients examined by the WECAC technique compared with the WE technique. The analysis will compare 2 groups (WECAC vs. WE) with a two sample t-test. If the data show substantial non-normality (as in a prior dataset with WE (5)); we will compare the groups with a Mann-Whitney *U* test; $P<0.0492$ considered significant. Secondary analysis of the primary outcome will utilize mixed effects regression models to evaluate the effects of potential covariates/confounds described in Tables 9 and 10. First, we will assess the effect of each individual covariate in regression models including the effect for group assignment and gender (stratification factor in the randomization), as well as the covariate by group interaction. Next, we will construct a multiple regression model with the set of covariates found to be significant (either as a main effect or an interaction effect) in an overall regression model along with group assignment.

Dependent Variables	Variable Type	Analysis method
Primary Outcome		
RTMIP (0=none, 10=most severe)	Continuous	2 sample Student t test. However, if distribution of the data is skewed as seen in an earlier study (5) , the Mann-Whitney *U* test will be used.

We have updated the analytic plan to include multi-level random effects (hierarchical model) for study site and provider within site. We will include two levels of random effects in this model: 1. random effects for study site, 2. random effects for provider which will be nested within sites.

Secondary hypotheses. *WECAC produces significantly higher quality colonoscopy outcomes. For example, when WE alone and WECAC are compared, the proportion completing colonoscopy with no pain, insertion time, and ADR in the proximal colon will be significantly different. For continuous secondary outcomes (e.g. insertion time), we will use a similar analysis strategy as that used for the primary out-outcome. For dichotomous secondary outcomes (e.g., no pain, proximal colon ADR) we will use Chi-square tests to compare the outcomes between the groups. Generalized logistic models will be used to adjust for covariates, gender and study site for these outcomes.*

Dependent Variables	Variable Type	Analysis method
Insertion time	*Continuous*	*Student t test*
No insertion pain (yes, no); proximal colon adenoma detection (yes, no)	*Binary*	*Chi square*

Exploratory hypotheses. Co-variables affect patients' report of RTMIP. Patient-related factors include age, gender, ethnicity, education, BMI, indication for colonoscopy, current symptomatic hemorrhoids,

Dependent Variables	Variable	Analysis method
Primary Outcome		
RTMIP	Continuous	Mixed effects regression, predict lower insertion pain, predictors-covariables

Co-variables	Variable	Analysis method
Recalled maximum insertion pain score (blinded) [validates primary]	*Continuous*	*Correlate RTMIP*
Body mass index (yes, no)	Continuous	
Age (<40 years) (yes, no); gender (male, female); ethnicity (White, Black, Hispanic, Asian, other) (yes, no); indication (Screening, surveillance, diagnostic); hemorrhoids (yes, no); chronic pain treatments (yes, no); substance abuse (yes, no); prior abdominal surgery (yes, no), sedation colonoscopy (yes, no), unsedated colonoscopy (yes, no), painful colonoscopy (yes, no); diverticulitis (yes, no); indication (abdominal pain) (yes, no); anticipation of pain (yes, no);	Binary	
Exploratory variables		
Staff satisfaction (0=not satisfied to 10=very satisfied)	Continuous	
Adequacy of patient blinding (yes, no); Loop reduction maneuvers (yes, no); Position change (yes, no); Abdominal compression (yes, no); Need to change technique (yes, no); Failed (poor prep) (yes, no); Pain tolerance (patient reported) (0=low, 1=high); Reason for unsedated (no ride, personal preference, other) (yes, no); Previous endoscopy experience (0=not good, 1=good); Objective anxiety (0=none, 1=high) (by endoscopist before examination); overall adenoma detection - entire colon (yes, no);	Binary	

chronic pain treatment, substance abuse, prior abdominal surgery, prior colonoscopy with or without sedation, current abdominal discomfort, expectation on day of colonoscopy, patient reported pain tolerance and subjective anxiety, objective anxiety assessed by colonoscopist, and reason for choice of no sedation. Procedure-related factors include technique of insertion, use of loop reduction maneuvers by the colonoscopist, change of patient position, application of abdominal compression by the assistant and insertion time, and length of colonoscope needed to reach the cecum. Examples: patients are more likely to report higher RTMIP amongst those with younger age, longer duration of insertion, lower expectation of painless insertion, female gender, low BMI, more frequent need for position change or abdominal compression. More patients will report lower RTMIP if there is low patient anxiety, good expectation of today's examination, no

prior abdominal surgery, colonoscopist use loop reduction maneuvers. The analysis of the exploratory outcomes will follow the same structure as the primary and secondary outcomes. To control for the multiplicity of exploratory outcomes we use the false-discovery rate (FDR) of Benjamini and Hochberg (67) rather than the traditional Bonferroni adjustment. The FDR assesses estimates the proportion of significant findings that are false positive results. The FDR method is less conservative and better preserves our ability to detect differences between groups.

Interim analysis. After 50% of subjects complete the protocol and data collection we will perform an inferential analysis of the primary endpoint as described in the statistical methods section. The only difference between the interim analysis of the primary endpoint and the final analysis will be the level of significance of the tests. If there is a significant treatment effect, this would communicated to the study Data and Safety Monitoring Board (DSMB) to consider early termination for efficacy.

Safety and Monitoring. This study has **MODERATE** risk with the potential of uncommon serious adverse events or unanticipated problems involving risks to participants. The local site investigator will be responsible for following adverse event (AE) reporting requirements: a. Reviewing the accuracy and completeness of all AEs reported; b. Complying with study policies as well as local IRB policies for reporting adverse medical events (AMEs) and/or serious adverse events (SAEs); c. Reporting to the IRB safety issues reported to the site by the sponsor (PI), and; d. Closely monitoring research volunteers for any new AMEs or SAEs. An AME is defined as "any untoward medical occurrence in a patient or clinical investigation subject administered a pharmacological product which does not necessarily have to have a causal relationship with this treatment". An AME can therefore be any unfavorable or unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the study intervention. All AMEs with a reasonable causal relationship to the investigative treatment will be considered at least "possibly related". A definite relationship does not need to be established. SAEs are a subset of AMEs defined by the International Conference on Harmonization for Clinical Safety Data Management as untoward medical occurrences that at any dose: a. Result in death; b. Are life-threatening; c. Require inpatient hospitalization or prolongation of existing hospitalization; d. Result in persistent or significant disability or incapacity; e. Result in a congenital anomaly/birth defect; or f. Result in any other condition that, based upon medical judgment, may jeopardize the subject and require medical or surgical treatment to prevent one of the above outcomes.

AEs and SAEs. These will be monitored during colonoscopy; within 7 days; between 7 and 30 days.

For each type of AE and SAE, rates of occurrence will be compared between treatment groups. Logistic regression will be performed to assess whether specific patient level characteristics are associated with any of the most frequently AEs or SAEs, such as cardiopulmonary events, bleeding, and so on.

Data security. All data will be de-identified and appropriate and applicable VA data security measures will be implemented.

Site performance. The data monitoring committee evaluates recruitment and retention performance monthly. Other performance problems such as protocol deviations, poor data quality, missing or overdue data, and reasons for withdrawal from study will be tracked. A monthly study conference call for study personnel including PD/PI, and all sites will be held to review study recruitment, data quality and protocol adherence. Study sites will be put on probation for poor performance, including under-recruitment. Typically, the probationary period is three months, at which time the site may be taken off probation, have probation continued or have funding stopped, depending on their performance during the probationary period.

Potential problems, limitations of the proposed methods and alternative procedures.

1. *Presumed lack of power to see any differences in female subjects The actual target sample size of female Veterans will be 96 based on updated pilot data. Additional unpublished pilot data showed that of 400 consecutive colonoscopies performed by the PI, 17 were in females, of whom 4 (1% of 400) requested no sedation (personal preference or no escort). I'll assume 1/2 of 1% (0.5%) of all colonoscopies will be in scheduled unsedated female Veterans when the option is presented up front. In 51 month (4.25 year), 191 (0.5% of 38250) will choose the option and available for recruitment. A recruitment rate of 50% (96/191) will be needed. We observed that when female Veterans lacked knowledge of unsedated colonoscopy, and if the option was not offered, 100% would be scheduled sedated and asked have escorts; those without escorts were left out (in a recent 5 month audit conducted by the PI at VAGLAHS, 6 of 51 scheduled female Veterans had last minute cancellations due to no escorts). These missed opportunities will be a source that the research coordinators will focus on. At each site we'll request VA Women's Health Clinics to assist the coordinators, to*

inform female Veterans of the scheduled unsedated option. As pointed out by Women's Health Directors, VA statistics indicate that nationwide, female Veterans, 11.6% of all Veterans in 2013, are projected to increase to 13% in 2020. With post-deployment chronic diarrhea, abdominal pain, 35% have digestive system diagnosis; and after exclusion of Giardia and Amoebiasis, colonoscopy is frequently used in further assessment. The increase will further enrich the pool of available female Veterans. If 6 female Veterans in 5 months at one site are limited by no escorts, the total number during the study period will be 6/5 months x 51 months x 3 sites = 184; they will be targeted by the research coordinators for enrollment into the study. The concordance in the number (191 and 184) of available female Veterans based on estimates using two different approaches is remarkable and will enhance our confidence in recruitment. Since there will be a sufficient number of female Veterans available for recruitment, we'll drop other approaches that will require additional expenses to recruit non Veteran females (e.g., radio advertising).

2. Changes from sedated to unsedated and air to WE require deviation from the norm, and carry a high risk that the study results would not lead to a change in future practice. The PI has shown that when scheduled unsedated colonoscopy was introduced, 76% (increase from 36%) of colleagues at the West LA VAMC are willing to perform unsedated colonoscopy, and nearly half have taken up use of WE (15). We anticipate demonstration of a less painful technique will improve acceptance of the novel approaches.

3. The cap adds an extra cost. The proposed RCT will seek data to indicate the added cost is justifiable, if WECAC is indeed less painful than WE in the unsedated Veterans.

4. The study should include strict control groups of air or CO_2 to aid insertion. We have shown that WE produces significantly less insertion pain than air or CO_2 insufflation during insertion (4). The use air or CO_2 to aid insertion in control groups cannot be justified in the unsedated Veterans in the current protocol.

5. Local VAMC provided transportation will obviate the need to provide scheduled unsedated colonoscopy. The approach is not appropriate for Veterans with a preference for no sedation.

6. Development of enhanced service to Veterans is more appropriately grouped under health service research. The proposal does focus on techniques directly implemented by the colonoscopist and is therefore a clinical endoscopic service. Clinical research funding is therefore appropriate.

7. The colonoscopist and nurse cannot be blinded to treatment allocation and the primary outcome of RTMIP. The measure has been repeatedly validated (4,5,32). Randomly selected (determined by the statistician) coded digital records will be reviewed by an independent observer to confirm agreement of findings, success of cecal intubation, time taken (endoscopic clock) for insertion to the cecum. Other types of self-selection bias are more difficult to control by blinding. The ongoing results will not be disclosed to the endoscopists, who may change the approach if it is known that one group is doing better or worse than the other. All investigators will be blinded to study results until completion of data collection.

8. Investigators may alter the WE technique (e.g., more frequent shortening maneuver) making it more difficult to detect a difference between techniques. Since the adjunct maneuvers are standard, investigators will be asked to use them consistently, and their use will be recorded. The RCT will be carried out by experienced colonoscopists with varying exposures to WE and WECAC, to assess the generalizability of the findings. The insertion technique for the water groups will be standardized by having the air pump turned off until the cecum is reached. Investigators will adhere to the details of both techniques described in the Methods section. There will be no limit on the water volume. Close attention will be paid to keep record of these variables to detect variations that may impact on interpretation of the results.

9. Differential dropout can result from patients being more likely to drop out of the trial if they have pain during the examination; resulting in more people dropping out of the control arm compared with the active treatment arm (presumably with less pain). They will be retained in the data analysis (intention-to-treat) and labeled treatment failures to avoid the differential dropout rate seriously biasing outcomes.

10. A different problem can occur in an intervention study if the subjects are not randomized to the treatment group they prefer: the subjects may drop out of the study before beginning treatment. In this study patients will be blinded to treatment assignment.

11. By the time evaluation of the WECAC is completed new technologies will have evolved to make WECAC obsolete. WE has been well-described and convincingly reduces RTMIP. The WECAC technique has been reported by us. Caps with other configuration (Daisy) are being reported. Endo Ring, G-Eye colonoscope, Endocuff and Daisy Caps have not been evaluated for their impact on colonoscopy insertion pain.