Study Protocol and Statistical Analysis Plan

STUDY PROTOCOL

Behavioral Economic Applications to Geriatrics Leveraging Electronic Health Records (BEAGLE)

Principal Investigator:

Stephen D. Persell, MD, Associate Professor, Northwestern University Feinberg School of Medicine

Supported by:

The National Institute on Aging

[R33AG057383]

Version 2

Version Date: June 1, 2023

Clinicaltrials.gov identifier: NCT04289753

TABLE OF CONTENTS

STUDY TEAM ROSTER

Stephen D. Persell, **MD, MPH** (Principal Investigator), Division of General Internal Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, 750 N Lake Shore Drive, 10th floor, Chicago, IL 60611, Phone: 312-503-6464, Email: spersell@nm.org

Theresa Rowe, DO (Co-investigator),

Lucia C. Petito, PhD (Statistician), Division of Biostatistics, Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, 680 N. Lake Shore Drive, Suite 1400, Chicago, IL, 60611 Email: lucia.petito@northwestern.edu

Jody D. Ciolino, PhD (Statistician), Division of Biostatistics, Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, 680 N. Lake Shore Drive, Suite 1400, Chicago, IL, 60611 Email: jody.ciolino@northwestern.edu

Craig Fox, PhD (Co-investigator), Anderson School of Management, University of California at Los Angeles, Los Angeles, CA, USA. Email: craig.fox@anderson.ucla.edu

Noah Goldstein, PhD (Co-investigator), Anderson School of Management, University of California at Los Angeles, Los Angeles, CA, USA. Email: noah.goldstein@anderson.ucla.edu

Jeffrey Linder, MD, MPH (Co-investigator), Division of General Internal Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, 750 N Lake Shore Drive, 10th floor, Chicago, IL 60611, Email: Jlinder@northwestern.edu

Jason Doctor, PhD (Co-investigator), Schaeffer Center for Health Economics and Policy, University of Southern California, Los Angeles, CA, USA. Email: jdoctor@usc.edu

Daniella Meeker, PhD (Co-investigator),

Ji Young Lee, MS (Data Analyst), Division of General Internal Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, 750 N Lake Shore Drive, 10th floor, Chicago, IL 60611, Email: jlee1@northwestern.edu

Tiffany Brown, MPH (Project Manager), Division of General Internal Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, 750 N Lake Shore Drive, 10th floor, Chicago, IL 60611, Phone: 312-503-9068, Email: t-brown@northwestern.edu

Yaw A. Peprah, MPH (Project Coordinator), Division of General Internal Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, 750 N Lake Shore Drive, 10th floor, Chicago, IL 60611, Email: yaw.peprah@northwestern.eduFax for all members of study team: 312-503-2755

PARTICIPATING STUDY SITES

Northwestern University. Stephen D. Persell, MD, MPH, Division of General Interna Medicine, Feinberg School of Medicine, Northwestern University, 750 N Lake Shore Drive, 10th Floor, Chicago, IL 60611, spersell@nm.org

This study occurs within the Northwestern Medicine health care system.

PRÉCIS

Study Title

Behavioral Economic Applications to Geriatrics Leveraging Electronic Health Records (BEAGLE)

Objectives

The main intent of this study is to determine whether an intervention comprised of 3 electronic health record (EHR) nudges designed using behavioral economic concepts reduces rates of over-prescribing in 3 clinical areas: prostate screenings, urinary testing for non-specific reasons female patients, and treatment for diabetes mellitus (DM). Our primary hypothesis is that practices randomized to receive these nudges will have lower over-prescribing rates in these 3 domains compared to control practices.

Design and Outcomes

We will conduct a multi-site cluster randomized trial of a behavioral intervention targeting overuse of prostate screening, urinary testing, and treatment for diabetes in geriatric patients, with practice as the unit of randomization.

The 3 co-primary outcomes are measured at the patient level, and capture whether an eligible patient received unnecessary treatment for a particular measure. We will report marginal monthly rates of each outcome. These include:

1. Ordering of prostate specific antigen (PSA) for prostate cancer screening in adult men aged 76 years and older
2. Ordering of urinalysis or urine cultures (UA/UC) for non-specific reasons to identify bacteriuria in women aged 65 years and older
3. Overtreatment of diabetes (DM) with insulin or oral hypoglycemic drugs in adults aged 75 years and older.

In the main analysis, we will compare a 12-month historical period to an 18-month intervention period. In subsequent secondary analyses to determine persistence of study effects, we will track all 3 co-primary outcomes for another 12 months after the intervention ceases.

Interventions and Duration

We will randomize clinics to either a brief education arm (clinicians at those clinics receive an invitation to an online education module) or brief education plus 18 months of clinical decision support within the EHR for the three clinical topics above.

The intervention period will be 18 months long, followed by an additional 12 months of data collection to assess persistence of study effects. We will also use data from a 12 month historical baseline period immediately prior to the start of the intervention period.

Sample Size and Population

We will recruit clinicians (physicians, advance practice nurses, and physician assistants) from 60 primary care clinics (internal medicine, family medicine and geriatrics) affiliated with Northwestern Medicine. Approximately 350 eligible clinicians will be recruited for this study.

We will randomize practices to intervention or control (1:1 allocation) using a constrained randomization algorithm, to achieve balance between study arms on baseline clinic characteristics.

1. STUDY OBJECTIVES

1.1 Primary Objective

Here, our primary aim is to evaluate the effects of an electronic health record (EHR) nudge designed to reduce over testing and treatment in geriatric patients for 3 clinical issues using a pragmatic cluster randomized controlled trial. We will measure the impact on validated quality measures associated with each of the 3 types of 'nudge interventions' and indicators of patient safety. The 3 clinical areas targeted by this nudge are:

1) Ordering of prostate specific antigen (PSA) for prostate cancer screening in adult men aged 76 years and older
(2) Ordering of urinalysis or urine cultures (UA/UC) for non-specific reasons to identify bacteriuria in women aged 65 years and older
(3) Overtreatment of diabetes (DM) with insulin or oral hypoglycemic drugs in adults aged 75 years and older.

1.2 Secondary Objectives

1.2.2. We plan to evaluate the persistence of any effects observed in Aim 1 by extending the analysis to include another 12 months of follow-up data.

1.2.1 Using a clinician survey administered twice – once pre-intervention roll-out and once at the end of the intervention period, we will measure the impact of our intervention on clinician attitudes.

2. BACKGROUND AND RATIONALE

2.1 Background on Overtreatment and Unnecessary Testing in Geriatric Populations

Despite strong recommendations from the American Geriatrics Society as part of the Choosing Wisely Initiative, sometimes clinicians do not follow best practices regarding ordering testing and treatments for their older adults that may lead to patient harm. These include: (1) testing and treatment for asymptomatic bacteriuria (ASB) in older women, (2) prostate specific antigen (PSA) screening in older men without prostate cancer, and (3) aggressive treatment with insulin or an oral hypoglycemic for older patients with type 2 diabetes.

Clinical decision support nudges, informed by behavioral economics and social psychology and delivered via EHRs, are promising strategies to reduce the misuse of services. Behavioral economics-informed interventions influence conscious and unconscious drivers of clinical decision making, are low cost to implement and disseminate, and can be incorporated into existing delivery systems. Our team has successfully employed these methods to the overuse of antibiotics for acute respiratory infections and has the necessary expertise in primary care practice, geriatrics, health informatics, social psychology, quality improvement, pragmatic trials and performance measurement to make this project a success. As our population continues to age, effective strategies to improve clinical care of older adults by reducing misused testing (e.g., urinalysis for ASB, prostate cancer screening with PSA) and treatment (e.g., overly-intense treatment of diabetes with insulin or other oral hypoglycemic) are critically needed.

2.2. Study Rationale

Preliminary data. This study builds on a pilot study conducted at Northwestern University by the same study team, which was funded by the National Institute on Aging (R21 #____; clinicaltrials.gov identifier NCT03704389). The EHR nudge used here, presented as clinical decision support prompts, were designed and pilot tested as part of that study.

During 2017-2018, our team interviewed NM primary care physicians caring for older adult patients to identify potential drivers of overuse of testing and treatment. Clinician participants reported knowledge of the Choosing Wisely recommendations (e.g., "We've always known to be a little more lenient for elderly patients because of hypoglycemia and falls…"), a reliance on patient preference (e.g., "Some patients are very insistent or expectant that they would want to continue to get testing"), clinical uncertainty and an underestimation of harms (e.g., "Don't want to over test her, but at the same time if she has treatable UTI want to treat her to get her some relief."). Additionally, we found that clinicians often perceive discussions with patients about not ordering tests or treatments to be time consuming (e.g., "After long discussion, seems overwhelming majority of patients want to do the test. Given limits of time we have, I've been less detailed in discussion…"). Finally, the theme of clinical inertia or resistance to change was present (e.g., "If this patient is doing well, I might leave it [diabetes treatment] alone.")

In fall 2018, we used data from these interviews and worked with clinical and decision science experts to design a nudge comprised of three clinical decision support interventions. The study team extensively reviewed the language and programming decisions prior to launching a pilot test of these interventions in January 2019. We consented and enrolled fourteen Northwestern Medicine primary care physicians to receive clinical decision support within NM's Epic electronic health record (EHR) when conditions within their patient's chart met nudge trigger eligibility criteria. Images of the 3 best practice advisories (BPAs)tested in the pilot intervention are included as Protocol Supplement 1 (pg. 2-3). The pilot study ran for six months between January-July 2019. We did not receive any clinician complaints regarding the interventions during the study period. In June 2019, we fielded a clinician survey to collect their feedback on their experience with study nudges. We will use the three interventions in this randomized controlled trial. During the R21 pilot study, the PSA alert fired 15 times, in 2 cases (13%) the test was not done and in 3 (20%) cases the test was ordered and an exception added to the report. The UA/UC alerts fired 34 times and in 21 cases (62%) the test was not ordered. Finally, the DM alert fired 93 times and on chart review 34% of patients whose clinician received DM CDS had diabetes therapy reduced during the pilot period.

Rationale for Accountable Justification.
Asking clinicians to provide an accountable justification for their ordering action that is not consistent with professional guidelines or best practices may reframe the behavior as non-normative by requiring the clinician to supply in the EHR a justification that becomes part of the record when the clinical action is taken. In addition to changing perceptions of what is normative behavior, entering a note that will go into the patient record may engage feelings of social accountability that promote more careful reflection.

3. STUDY DESIGN

The Behavioral Economic Applications to Geriatrics Leveraging Electronic Health Records (BEAGLE) trial is a single site, cluster-randomized controlled trial with practice as the unit of randomization. The primary aim is to test the ability of an intervention based on behavioral economic principles to reduce the

rates of commonly occurring inappropriate testing and treatment in geriatric patients across 3 domains: PSA, UA/UC, and DM. We will randomize practices using a 1:1 allocation; clinicians at intervention practices will receive the intervention when an eligible patient is seen. If physicians at intervention clinics wish to override the clinical decision support tool, they will need to provide justification that appears in the patient's EHR (accountable justification). Clinicians at control practices will receive a one-time education module about the risks of over-treatment and testing in these areas.

We will enroll approximately 350 clinicians from 60 participating primary care clinic sites and collect baseline data for the 12 months leading up to the intervention start (monthly data collected using a look-back window of 12 months). All clinicians will receive a one-time education module about the risks of over-treatment and testing in these areas at the time of consent and enrollment. Clinician surveys will be administered at the time of enrollment and after the intervention is completed. The 3 co-primary outcomes are quality measures [detailed below in section 6]. Secondary outcomes are patient safety measures [detailed below in section 6]. The 18-month intervention period will be followed by a 12-month follow-up period to measure persistence of effects after interventions end. Patient-level data from EHRs at participating practices were abstracted at 3 points [detailed below in section 11].

4. SELECTION AND ENROLLMENT OF PARTICIPANTS

This clinical trial will be conducted at 60 clinics within Northwestern Medicine: Northwestern Medical Group in the North and Central regions and Regional Medical Group located in the western and northwestern Chicago suburbs. Leadership for these regions endorsed the system-wide conduct of this trial and its performance with a waiver of informed consent.

4.1 Clinic Eligibility

Practices that deliver outpatient primary care and include internal medicine, family medicine or geriatrics physicians or advanced practice providers will be included in this study. We will exclude practices that only have residents or only provide emergency care.

4.2 Patient eligibility

All patients aged 65 years and older will be included in the overall study patient population. We will assess study outcomes for all clinics, clinicians and individual patients who meet the pre-defined study eligibility criteria for each measure.

Co-Primary Outcome	Eligibility Criteria
PSA	All adult men aged 76 years and older with a visit to a participating primary care or geriatrics NM clinic between September 1, 2018 – February 28, 2023
DM	All adults aged ≥ 75 years old with a visit to a participating primary care or geriatrics NM clinic visit between September 1, 2018 – February 28, 2023
UA/UC	All adult women aged 65 years and older with at least one primary care, geriatrics, or immediate care NM visit between September 1, 2018 – February 28, 2023

Note: all subjects over the age of 89 years will be aggregated into a single category. We will obtain these outcomes by queries of the NMEDW. Some data are prospective in that they do not exist at the time of

IRB submission but retrospective in that we will only have access to them after collected in NM Epic as part of routine care delivery.

5. STUDY INTERVENTIONS

5.1 Definitions of Study Interventions

Clinician education (control). An interactive educational module (15-20 minutes) covering the three targeted areas of overuse.

Clinical decision support plus clinician education (intervention).

Best Practice Advisories were programmed into Epic to present content to the clinician and feature functionality to add content to the encounter documentation. Clinicians attributed to clinics in the 'CDS intervention + Education' arm receive CDS alerts when conditions within a chart meet triggering criteria. For example, a CDS alert appears when a clinician places an order for a PSA test when the patient meets qualifying criteria (e.g., ≥ 76 years old, no prior prostate cancer). The CDS alerts encourage the clinician to cancel the order in the cases of PSA and urine testing or to decrease diabetes medication in the case of diabetes overtreatment (Table 1); however, they do not restrict clinical decision making in any way. If the clinician proceeds with the order for the PSA, urinalysis, or urine culture they are asked to enter text justifying the order. This justification is then included within the encounter report, visible to any clinician within the chart, under a heading titled 'Testing Justification'. If the clinician does not enter a justification, the note "No justification given" is included in the encounter report.[9, 10] Images of the CDS presented to the clinicians for each alert and triggering criteria are included in the supplemental material. The EHR programming logic used in the pilot study was re-tested for this trial.

5.2 Administration and Duration of Interventions

The clinician educational module will be administered by the study team. The nudges will be delivered via NM's instance of Epic and administered by NM IT.

The intervention period will be 18 months in length for all participants. We will continue to collect data on all study outcomes for 12 months after the intervention ends to measure persistence of effects. The pre-intervention baseline period will be 12 months in length.

5.3 Interruption due to COVID-19 Pandemic

The intervention period for BEAGLE was originally slated to begin in April 2020. However, on March 21, 2020, the governor of Illinois and mayor of Chicago issued a "stay-at-home" order to mitigate harm from the COVID-19 pandemic. This severely impacted regular primary care activities at Northwestern Medicine. As such, we chose to delay the start of the intervention until regular care resumed. The intervention was rolled out on September 1, 2020, and the look-back period for the study measures was adjusted to exclude data collected between March 1, 2020 and August 31, 2020. So, for example, study outcomes reported in December 2020 included the following 12 months of data: June 1, 2019 – February 29, 2020 plus September 1, 2020 – November 30, 2020.

6. STUDY ENDPOINTS

6.1 Primary endpoints

Our team developed and validated by manual chart review three electronic clinical quality measures (one associated with each of the 3 study interventions). These measures will be abstracted from the EHR once a month for the duration of the intervention period. The three co-primary endpoints are:

PSA. The PSA screening in older men (PSA) measure is defined as the presence of a PSA laboratory result in the EHR during the measurement period among men aged 76 years and older at the start of the measurement period with at least one visit during the measurement period with a clinician included in the study and who are without a diagnosis or procedure suggesting a history of prostate cancer (Protocol Supplement 2, pg. 4). Patients on androgenic agents or anabolic steroids are excluded (Protocol Supplement 2, pg. 7). Labs constituting the numerator are provided in Protocol Supplement 2 (pg. 7). Patients are attributed to the primary care clinician with whom they have the greatest number of visits during the measurement period. The measurement period consists of the 365 days preceding the measurement date.

UA/UC. Urine Testing for Non-Specific Reasons (UA/UC) is defined as the absence of a diagnostic code for a specific genitourinary sign, symptoms or other potentially relevant indication (Protocol Supplement 3, pg. 8) among instances where a urinalysis (UA) and/or urine culture (UC) is obtained in the interval 24 hours before to 48 hours after a face-to-face ambulatory care visit by a woman aged 65 years or older with a qualifying clinician during the study period. Testing episodes are attributed to the clinician who ordered the test. Only UA and/or UC ordered by a qualifying clinician are included. Only the first testing episode in a measurement period is included in patients with multiple qualifying visits. The measurement period is the 365 days preceding the measurement date.

DM. Diabetes Overtreatment in the Elderly (DM) is defined as the having the most recent hemoglobin A1C during the measurement period or during the year prior to the measurement period (if no A1C obtained during the measurement period) of less than 7.0 among adults aged 75 years and older with a diagnosis of diabetes mellitus who have insulin or an oral hypoglycemic (e.g., sulfonylurea or meglitinide, Protocol Supplement 4, pg. 12) on their active medication list at the end of the study period. Patients are attributed to the primary care clinician with whom they had the greatest number of visits during the measurement period. The measurement period is the 365 days preceding the measurement date.

6.2 Secondary endpoints

We chose secondary endpoints related to the primary PSA, UA/UC, and DM outcomes. These were selected to identify other burdens on the healthcare system, and included:

- PSA
    - Prostate biopsy in older men
    - Prostate MRI in older men
    - New prostate cancer diagnosis in older men
- UA/UC
    - Total urinalysis use in women aged 65 and older
    - Total urine culture use in women age 65 and older
    - Antibiotic treatment in women age 65 and older (all oral UTI treatments)
    - Antibiotic treatment in women age 65 and older (nitrofurantoin and Fosfomycin only)
    - C. difficile infection in women age 65 and older

    - Rash in women age 65 and older
- DM
    - Hypoglycemia requiring an urgent care visit (not emergency department) among eligible patients

We will also measure the three co-primary outcomes monthly during a follow-up period (months 19-30) to evaluate the persistence of any effects observed during the study period.

We will survey the clinicians who consent to participate in an online questionnaire about their attitudes about their experience with study interventions as well as their attitudes and beliefs about testing and treatment more broadly among geriatrics patients. See Baseline Clinician Survey and Follow-up Clinician Survey.

7. STUDY PROCEDURES

7.1 Schedule of Evaluations

Assessments	Historical practice: Baseline overuse (Month -17 to Month 0)	Baseline, Enrollment, Randomization (Month 0, Day 1)	Intervention Start (Month 1)	Monitored and Measured Monthly	Intervention End (Month 18)	Follow-up Period (Month 19 to Month 30)
Clinician-level Assessments						
Informed Consent		X				
Demographics		X				
Inclusion/Exclusion	X					
Provider Attitudes Survey		X			X	
Patient-level Assessments						
ICD-10 codes	X	X	X	X	X	X
Ordering	X	X	X	X	X	X
Adverse Events	X		X	X	X	

7.2 Description of Evaluations

7.2.1 Consent. We will send providers at participating sites an introductory email that includes a description of the broad goals of the study, a general description of the intervention, and a link to the electronic consent form and baseline survey. The consent document will indicate that participation is voluntary and that decisions to participate (or not) will have no bearing on any provider's status at his or her clinic. Providers who provide consent to participate will be asked to complete an online survey and brief educational session prior to the intervention phase, permit de-identified patient records pertaining to eligible patients who saw them to be included in the study database, and complete a 15-minute post-intervention survey. We will also describe compensation that providers will receive for participation.

7.2.2 Enrollment. The enrollment date will be documented on the online consent form at the time of consent. The interventions will be activated after all clinicians have been enrolled or declined to participate.

7.2.3. Baseline assessments. We will collect information about the rates of each of the 3 co-primary outcomes monthly for the 12 months leading up to the activation of the intervention. This period will document historical practice. We will also administer a survey to assess provider characteristics and provider attitudes toward practice guidelines, clinical decision support, and practice environment.

8. SAFETY ASSESSMENTS

We did not study any safety endpoints related to the PSA primary outcome. The safety endpoints related to the UA/UC and DM primary outcomes were deemed serious adverse events, and included:

- EHR-identified ED or hospital care possibly due to UTI or sepsis among women 65 and over following an office visit
- EHR-Identified ED or hospital care possibly due to hyperglycemia among previously tightly controlled
- Poor diabetes control (HbA1c>9%) among individuals who were previously tightly controlled (HbA1c < 7%)

The Principal Investigator will report any unanticipated events to the IRB as well as the Data Safety and Monitoring Board (DSMB) assembled for this study. When notified of an unanticipated event, the DSMB will convene and make a decision as to whether the study should continue. The IRB will also be notified of the DSMB's decision.

8.1 Safety Monitoring and Reporting Procedures

The Principal Investigator will report any unanticipated events to the IRB as well as the Data Safety and Monitoring Board (DSMB) assembled for this study (listed in Section 14 below). When notified of an unanticipated event, the DSMB will convene and make a decision as to whether the study should continue. The IRB will also be notified of the DSMB's decision.

9. INTERVENTION DISCONTINUATION

Following each DSMB meeting, the board will make recommendations to the local IRBs as to whether the study should continue or if changes to the protocol are necessary for continuation.

10. STATISTICAL CONSIDERATIONS

10.1 General Design Overview

Hypotheses

Our primary hypothesis is that practices randomized to receive behavioral economic interventions will have decreases over time in the over-screening, over-treatment, and over-testing of older patients who

may no longer warrant such care. This hypothesis will be evaluated in an intention-to-treat difference-in-differences framework using a mixed-effects logistic regression model. Fixed effects will include the effects of interventions over time (i.e., interactions between randomization assignment and time), using a historical period 12 months prior to the intervention period. Providers and randomization unit (clinic) will be modeled as random effects.

Design

We will conduct a cluster randomized trial using data abstracted from the EHR at one academic health system. We chose a cluster-randomized design at the clinic level to minimize spillover effects – contamination that might occur if individual providers in close proximity are randomized to different interventions. The study design for BEAGLE included the implementation of a modified constrained randomization scheme, to ensure that clinics in each study arm were similar prior to activation of the intervention.

Outcome measures

This study has 3 co-primary outcomes, referred to as PSA, UA/UC, and DM (defined in Section 6 above, and further detailed in Section 10.4 below). They capture the over-testing, over-screening, and over-treatment of older patients, whose medical needs may have evolved to require less care.

10.2 Sample Size and Randomization

Power Calculations

Study power for BEAGLE was designed around estimating minimum detectable absolute rate reductions for the three co-primary outcome measures at the individual level. Power calculations were done for a 2-level mixed effects logistic regression model that included a random intercept for clinic using PASS software. Historic data, from a period spanning X, 2017 – Y, 2018 were used to estimate an inter-clinic correlation coefficient of 0.017. In all calculations, we assumed a two-sided type I error of 5% (Bonferrroni-corrected to 0.05/3 = 0.017), 80% power, 30 clinics in each arm, and that patients would be divided equally between clinics in 2 groups. Initial calculations assumed baseline event rates and sample sizes described in Table 1. These analyses should have been conservative, as our assumed sample size was derived from an historic 12-month period, while the study follow-up period was 18 months. We did not assume any clinic drop-out, as all data will be available via the EHR. With these assumptions, we anticipated having adequate power to detect an absolute difference of roughly 4% for each of the measures. For the PSA and UA/UC outcomes, this corresponded to a relative risk between 0.80 and 0.87; for the DM outcome, it corresponded to a relative risk between 0.73 and 0.81.

Table 1. Summary of power calculations for the BEAGLE study.

	PSA N=9,500			**DM N = 6,600**			**UA/UC N = 15,000**		
Baseline event rate	20%	25%	30%	15%	20%	25%	20%	25%	30%
Minimum detectable absolute rate reduction	4.1%	4.5%	4.8%	3.9%	4.4%	4.8%	3.8%	4.2%	4.4%
Corresponding relative risk	0.80	0.82	0.84	0.74	0.78	0.81	0.81	0.83	0.85

Treatment Assignment Procedures – Constrained Randomization

Clinics were randomized to receive either the active intervention or control. Clinicians were attributed to precisely one clinic based on the clinic where they spent the plurality of their time in March 2019 – February 2020. All 60 eligible primary care clinics seeing adult patients were randomized. The study design for BEAGLE implemented a modified constrained randomization scheme. Eight clinic-level variables were selected based on their possible influence on the three co-primary outcomes, and were incorporated into the randomization schemes in a principled manner to reduce potential bias and promote comparable arms at the analytic unit (clinician/patient encounter). These clinic-specific variables, created using data abstracted from the EHR, were:

1. Health system region – We categorized each practice into one of four geographic areas in the Chicago area (categorical).
2. Number of clinicians attributed to each practice (continuous).
3. Number of men eligible for PSA overtesting (age 76 years or older, no history of prostate cancer, not on androgenic agents or anabolic steroids) seen in clinic between March 1, 2019 and February 29, 2020 by a participating clinician (continuous).
4. Annual PSA overtesting rate (percentage).
5. Number of women eligible for unnecessary UA/UC testing measure between March 1, 2019 and February 29, 2020 (continuous).
6. Annual UA/UC overtesting rate (percentage).
7. Number of patients eligible for DM overtreatment measure between March 1, 2019 and February 29, 2020 (continuous).
8. Annual DM overtreatment rate (percentage).

Six practices only contributed nine months of data (June 1, 2019 – February 29, 2020) because they were acquired by Northwestern Medicine between March 1, 2019 and June 1, 2019.

We used the methodology presented in [cite] to implement the constrained randomization algorithm. First, the study biostatistician (LCP) generated possible 1:1 randomization schemes. Since there were infinitely many ways to randomize the 60 practices into 2 groups, the study biostatistician decided to create 100,000 possible sequences. Second, the balance of the eight prespecified variables (which were "constrained" upon) was then evaluated in the 100,000 possibilities. Schemes were deemed candidate schemes if they had "adequate balance" – that is, the measure of imbalance for each variable (separately) could not surpass a pre-specified threshold. For categorical variables, we allowed no more than a difference of two practices across each arm. For continuous variables, we compared sample means across study arm, and adapted the ideas of the minimal sufficient balance approach. We called a variable 'adequately balanced' if a T-statistic from an independent two sample t-test with equal variance was between -0.385 and 0.385, corresponding to a p-value for the individual variable of 0.30. Last, the final sequence was selected from the candidate schemes (which were adequately balanced) at random with equal probability.

As was disclosed in the study protocol paper, we achieved adequate balance for seven of the eight pre-specified variables used in the constrained randomization algorithm. Due to a coding error, we applied balancing criteria for the PSA measure on the number of eligible patients (denominator) and the number of PSA orders (numerator) rather than the annual PSA overtesting rate (#4 above). This is an absolute measure instead of a relative measure that accounts for the size of the eligible patient population at each clinic. For a complete discussion of the implications of this error, please reference the protocol paper.

The study biostatistician (LCP) generated the final randomization in R version 3.6.4 (R Core Team, 2020). [25] Randomization occurred in June 2020. The allocation of the sequence was concealed until

after the intervention period began in September 2020 (due to interruption of regular care during the COVID-19 pandemic).

10.3 Interim analyses and stopping rules

No interim analyses were conducted for primary or secondary outcomes. The Data Safety and Monitoring board is granted the power to recommend discontinuation of the study to the IRB, if safety concerns are found. The DSMB will meet at least three times over the course of the study: prior to the commencement of the study, at approximately month 9 of the intervention period, and at the end of the intervention period. Additional meetings will occur as needed or as directed by the study sponsor. Following each meeting, the board will make recommendations to the IRB as to whether the study should continue or if changes to the protocol are needed.

10.4 Outcomes

10.4.1 Co-primary outcomes

This study has 3 co-primary outcomes, referred to as PSA, UA/UC, and DM (defined in Section 6 above). They capture the over-testing, over-screening, and over-treatment of older patients, who may need less care than they are receiving. The ICD-10 codes used to capture these co-primary outcomes are defined in detail in the Supplement to this document. These outcomes are computable clinical quality measures that are abstracted from the EHR. They were previously validated as part of pilot work for this study. These will be extracted from the EHR on a monthly basis, with data for each month representing a 365 look-back period.

An office visit will be eligible for inclusion in the outcome denominator if it meets the following criteria:

PSA. Men were eligible for the PSA over-screening measure if they were age 76 years or older and attended at least one visit in the 365 days prior to the measurement date (for example, 365 days prior to February 1, 2020) with a clinician included in the study. Eligible patients had no history of prostate cancer (determined via diagnosis or procedure code) and were not taking an androgenic agent or anabolic steroid as of the measurement date.

UA/UC. Women were eligible for the urine testing for non-specific reasons (UA/UC) measure if they were age 65 years or older and had an in-person ambulatory care visit with a participating clinician in the 365 days prior to the measurement date where a urinalysis and/or urine culture was obtained in the interval 24 hours before to 48 hours after the visit. If a woman had multiple qualifying visits in the 365-day window, only the first was included.

DM. Patients were eligible for the diabetes over-treatment measure if they were age 75 years or older with a diagnosis of diabetes mellitus and their most recent hemoglobin A1C in the 365 days prior to the measurement date was less than 7.0%. Patients were attributed to the primary care clinician with whom they had the greatest number of visits in the 365-day window.

10.4.2. Secondary and safety outcomes

The secondary outcomes related to the primary PSA, UA/UC, and DM outcomes included:

- PSA
    - Prostate biopsy in older men
    - Prostate MRI in older men
    - New prostate cancer diagnosis in older men

- UA/UC
    - Total urinalysis use in women aged 65 and older
    - Total urine culture use in women age 65 and older
    - Antibiotic treatment in women age 65 and older (all oral UTI treatments)
    - Antibiotic treatment in women age 65 and older (nitrofurantoin and Fosfomycin only)
    - C. difficile infection in women age 65 and older
    - Rash in women age 65 and older
- DM
    - Hypoglycemia requiring an urgent care visit (not emergency department) among eligible patients

We will also measure the three co-primary outcomes monthly during a follow-up period (months 19-30) to evaluate the persistence of any effects observed during the study period.

We did not study any safety endpoints related to the PSA primary outcome. The safety endpoints related to the UA/UC and DM primary outcomes were deemed serious adverse events, and included:

- EHR-identified ED or hospital care possibly due to UTI or sepsis among women 65 and over following an office visit
- EHR-Identified ED or hospital care possibly due to hyperglycemia among previously tightly controlled
- Poor diabetes control (HbA1c>9%) among individuals who were previously tightly controlled (HbA1c < 7%)

10.5 Statistical Analyses

We will use the following descriptive statistics to characterize the sample: Means and medians for continuous measures, frequencies for count data, standard deviations and interquartile ranges for variance.

To test our 3 co-primary hypotheses, we will employ a mixed-effects hierarchical logistic regression model to estimate the adjusted marginal effect over time of each intervention on the 3 co-primary outcomes. Fixed effects will include intervention assignment, time period (historical baseline versus intervention period), time period interacted with intervention assignment, time (continuous, linear), and a 3-way interaction term between time period, intervention, and time (continuous). The coefficient on the 3-way interaction term will represent whether the log-odds of the monthly rate of change in each co-primary outcome differs between intervention and control patients during the intervention period. Providers will be included as random effects. To isolate the effect of the intervention on the co-primary outcomes, we will also include fixed effects for the variables included in the constrained randomization procedure.

The secondary and safety outcomes will be examined in a similar fashion. If they are sufficiently rare (for example, no events in a given month), model convergence may become an issue. First, we plan to expand the time frame from months to quarters and fit the same model. If that fails, we will abandon the time-varying aspect of the analysis and do one pull for the historical baseline period and one pull for the intervention period. The mixed effects logistic regression model would then be run without including continuous time. If that still fails to converge, we will exclude the random effects for providers and use a fixed effects logistic regression model.

Last, the three co-primary outcomes of interest will be examined in the following pre-specified subgroups:

- Sex (DM measures only)
- Race and ethnicity
- Age groups (for UA/UC: 65-74, 75-84, 85+) (for PSA: 76-79, 80-84, 85+) (for DM: 75-79, 80-84, 85+). We will also examine for effects on patients younger than the target popuations (for UA/UC: 50-64, PSA: 70-75, DM 65-74).

11. DATA COLLECTION AND QUALITY ASSURANCE

11.1 Data Collection Forms

Two types of data were collected: 1) data from electronic medical and billing records, abstracted from Northwestern Medicine's Electronic Data Warehouse; and 2) data from online surveys at the beginning and end of the study.

11.2 Data Management

The study analyst, JYL, extracted data from the NMEDW at 3 points: end of baseline period, end of intervention period, and end of additional follow-up period.

Both the baseline and 18M surveys will be administered via REDCap.

11.3 Quality Assurance

Extensive quality assurance was conducted during the pilot R21 phase of this study.

12. PARTICIPANT RIGHTS AND CONFIDENTIALITY

12.1 Institutional Review Board (IRB) Review

The study protocol and the informed consent document will be reviewed and approved by the Northwestern University Institutional Review Board (IRB).

12.2 Consent Process

Not applicable. We were granted a waiver of consent for clinician and patient record review. Here, we provide details about the alteration of the consent process for our online consent script ahead of baseline and study close-out clinician surveys.

12.2.1 Patient Consent Process

We were granted a waiver of the consent process to review (NMEDW query) patient medical records to collect study data because we met the following criteria:

- This research was NOT FDA-regulated and only involved adults
- The research involved no more than minimal risks to subjects. Review of medical record data had strict inclusion and exclusion criteria and identified data was only seen by the study analyst extracting data from EDW.
- This research could not be practicably be carried out without the waiver. We were attempting to describe clinician testing and treatment behavior across the entire geriatric patient population and to use these measures as the basis for clinician interventions that turn on and off based on data that reside within the EHR.
- To understand how the nudges work, we must have access to identified patient and clinician data. Our analyst assigned a study identifier and stripped out all identifiers prior to finalizing and sharing the study analytic files.
- The waiver will NOT adversely affect the rights and welfare of subjects.
- Subjects (patients) will not be provided any study information as they will be unaware of study participation.

12.2.2 Clinician Consent Process

We were also granted a waiver of consent for clinician record review because we met the following criteria:

- This research is not FDA-regulated. It is a quality improvement study designed to evaluate clinical decision supports impact on guideline recommended care. Clinicians' actions are in no way restricted.
- The research does not involve non-viable neonates. Adult NM clinicians will be presented alerts in the EHR when criteria within the chart are met.
- The research involves no more than minimal risk. Clinical decision supports are intended to help clinicians make safer clinical decisions for their patients. Their interactions with study nudges and their performance on study quality measures will have no effect on their NM employment. Performance data will be confidential.
- The waiver of consent will not adversely affect the rights and welfare of clinicians who interact with study nudges. Clinicians have choices in how they respond to nudges. Clinicians will be able to order tests and treatments in the exact same way as prior to the initiation of this pilot study.
- The research could not practicably be carried out without the waiver of consent. This study attempts to describe clinician interactions with EHR tools intended to improve quality of care. If we consented clinicians for this study, clinicians that would volunteer to participate would very likely be different from non-volunteering clinicians. *The goal of the study is to develop and test clinical decision supports that could be turned on for large populations of clinicians and not just the subgroup of clinicians that may be interested.*
- We will inform all participating clinicians about study nudges, will respond promptly to any questions or concerns they have about their experience interacting with them, and will provide summary quality measure data at the clinic level at the end of the study.

Further, we were granted a waiver of documentation of consent for the clinician surveys administered at baseline and end of the intervention. We used REDCap functionality to present IRB approved informed consent language as first section of survey and require a 'yes' response in order to proceed and view/interact with online survey. If a clinician refused participation by marking 'no' to consent item, they will only see a 'thank you, please close your browser' message and we no longer sent reminders to that clinician. Clinicians were instructed they can print the consent language from their browser if they would like a hard copy to keep for their records. The consent language was presented as first section on both the baseline and the 18M follow-up survey.

Both the study PI (SDP) and project manager (TB) had their contact details included with the online consent script. They responded promptly to any potential participant who reaches out with questions or concerns. The full study team reviewed the online consent language to ensure it had all informed consent elements that are necessary.

PROTECTED HEALTH INFORMATION (PHI AND HIPAA)

This study involves the use of Protected Personal Health Information (PHI) accessed through NM EDW. We were granted a waiver of HIPAA authorization. We have reviewed the HIPAA Waiver of Authorization checklist (HRP-441) and provide details below.

- We clearly define here the description of the PHI our study will access and why it is necessary for this research.
- The use of PHI involves no more than minimal risk to the privacy of individuals based on the following:
    - An adequate plan is in place to protect the identifiers from improper use and disclosure. See data security section above.
    - An adequate plan is in place to destroy the identifiers at the earliest opportunity consistent with conduct of the research. We will destroy identifiers as soon as study analytic files are created and study manuscripts accepted. Identifiers will be destroyed prior to the end of the grant period.
    - Adequate written assurances that the protected health information will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research study, or for other research for which the use or disclosure of protected health information for which an authorization or opportunity to agree or object is not required by 45 CFR 164.512.
- Given that the research describes the entire geriatric patient population with a visit during the study look-back or intervention period, this research could not practicably be conducted without the waiver.
- The research could not practicably be conducted without access to PHI.

12.3 Participant Confidentiality

We will only extract the minimum necessary data to answer research objectives. All analytic files will be de-identified.

Clinician surveys focus on their clinical role and attitudes and should not feel intrusive. However, clinicians can skip any survey items they wish and still submit the survey.

Clinic leaders and individual clinicians will be provided study team contact information so they can report concerns or complaints about study interventions at any time. We will also be informed of any tickets to Epic helpdesk that are related to study interventions by NM Epic programmers.

12.4 Study Discontinuation

While this was a minimal risk study, we established a DSMB. Following each DSMB meeting, the board made recommendations to the IRB as to whether the study should continue or if changes to the protocol were necessary for continuation.

13. COMMITTEES

The Data Safety Monitoring Board was comprised of:

- Jerry Gurwitz, MD, Chairperson
- Carl Pieper, DrPH, Biostatistician
- Michael Steinman, MD, Safety Officer

14. PUBLICATION OF RESEARCH FINDINGS

Publication of results from our research will follow the NIH Public Access Policy, which requires that we submit to the National Library of Medicine's PubMed Central an electronic version of final, peer-reviewed manuscripts upon acceptance for publication, to be made publicly available no later than 12 months after the official date of publication.

ABBREVIATIONS

ASB: asymptomatic bacteriuria
BEAGLE: Behavioral Economic Applications to Geriatrics Leveraging Electronic Health Records
BPA: best practice advisory
CDS: clinical decision support
DM: diabetes mellitus
DSMB: Data Safety and Monitoring Board
EHR: electronic health record
IRB: Institutional Review Board
NM: Northwestern Medicine
NMEDW: Northwestern Medicine Enterprise Data Warehouse
PSA: prostate specific antigen
UA/UC: urinalysis/urine culture