Contemporary Clinical Trials 112 (2022) 106649

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Design of Behavioral Economic Applications to Geriatrics Leveraging Electronic Health Records (BEAGLE): A pragmatic cluster randomized controlled trial

Tiffany Brown [a], Theresa A. Rowe [a], Ji Young Lee [a], Lucia C. Petito [b], Ryan Chmiel [c], Jody D. Ciolino [b], Jason N. Doctor [d], Craig R. Fox [e], Noah J. Goldstein [e], Darren Kaiser [c], Jeffrey A. Linder [a], Daniella Meeker [d], Yaw Peprah [a], Stephen D. Persell [a,f,*]

[a] Department of Medicine, Division of General Internal Medicine and Geriatrics, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[b] Department of Preventive Medicine, Division of Biostatistics, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[c] Northwestern Memorial HealthCare, Chicago, IL, USA
[d] Schaeffer Center for Health Economics and Policy, University of Southern California, Los Angeles, CA, USA
[e] Anderson School of Management, University of California at Los Angeles, Los Angeles, CA, USA
[f] Center for Primary Care Innovation, Institute for Public Health and Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

ARTICLE INFO

Keywords:
Overuse
Behavioral economics
Electronic health records
Prostate cancer screening
Diabetes mellitus
Urinary tract infections

ABSTRACT

Background: Overtesting and treatment of older patients is common and may lead to harms. The Choosing Wisely campaign has provided recommendations to reduce overtesting and overtreatment of older adults. Behavioral economics-informed interventions embedded within the electronic health record (EHR) have been shown to reduce overuse in several areas. Our objective is to conduct a parallel arm, pragmatic cluster-randomized trial to evaluate the effectiveness of behavioral-economics-informed clinical decision support (CDS) interventions previously piloted in primary care clinics and designed to reduce overtesting and overtreatment in older adults.

Methods/Design: This trial has two parallel arms: clinician education alone vs. clinician education plus behavioral-economics-informed CDS. There are three co-primary outcomes for this trial: (1) prostate-specific antigen (PSA) screening in older men, (2) urine testing for non-specific reasons in older women, and (3) overtreatment of diabetes in older adults. All eligible primary care clinics from a large regional health system were randomized using a modified constrained randomization process and their attributed clinicians were included. Clinicians were recruited to complete a survey and educational module. We randomized 60 primary care clinics with 374 primary care clinicians and achieved adequate balance between the study arms for prespecified constrained variables. Baseline annual overuse rates for the three co-primary outcomes were 25%, 23%, and 17% for the PSA, urine, and diabetes measures, respectively.

Discussion: This trial is evaluating behavioral-economics-informed EHR-embedded interventions to reduce overuse of specific tests and treatments for older adults. The study will evaluate the effectiveness and safety of these interventions.

1. Introduction

Overtesting and overtreatment are common in health care. [1–3] This may be particularly problematic for older adults leading to increased morbidity and mortality. [4] The Choosing Wisely campaign, an initiative of the American Board of Internal Medicine Foundation endorsed by the American Geriatric Society (AGS), has provided recommendations on reducing overtesting and overtreatment of older adults. [5] Three targets relevant to outpatient primary care are (1) do not use prostate specific antigen (PSA) screening for prostate cancer without considering life expectancy and the risks of testing, overdiagnosis, and overtreatment; [6] (2) do not use antimicrobials to treat bacteriuria in older adults unless specific urinary tract symptoms are present; [7] and (3) avoid using medications other than metformin to achieve hemoglobin A1C < 7.5% in most older adults with type 2 diabetes. [8] Overuse in these areas is common, and interventions are

* Corresponding author at: 750 N Lake Shore Dr., 10th floor, Chicago, IL 60611, USA.
E-mail address: spersell@nm.org (S.D. Persell).

https://doi.org/10.1016/j.cct.2021.106649
Received 8 September 2021; Received in revised form 1 December 2021; Accepted 4 December 2021
Available online 8 December 2021

needed to improve the quality of care for older adults.

Standard interventions grounded in the neoclassical model of rational decision making rely on education, simple financial incentives, and/or regulation to improve quality of care. Interventions guided by insights from behavioral economics and social psychology have shown promise in recent years, even in situations where standard interventions have failed. [9–13] Behavioral economics-informed interventions influence conscious and unconscious drivers of clinical decision making. One such driver may be that clinicians over-weight the low probability of identifying a treatment opportunity that could be discovered through testing that would lead to patient benefit (e.g., diagnosing cancer that would lead to harm during the patient's remaining lifetime). [14] This in turn may cause clinicians to fail to consider potential downsides of testing in specific populations. One behavioral approach that may decrease overtesting is to call attention to the potential harms of performing the clinical action which can make those harms (e.g., complications that can arise from testing) easier to imagine and receive greater weight. [15] Another approach might be to present social norm data showing that a large proportion of similar clinicians forego a particular test or procedure. [16] A third strategy, accountable justification interventions, prompting clinicians to provide a brief justification note when ordering tests or treatments that may not be guideline concordant, have been used to change clinician behavior. [9,17]

The Behavioral Economic Applications to Geriatrics Leveraging Electronic Health Records (BEAGLE) study aims to assess the ability of a set of CDS interventions informed by behavioral economics principles designed to reduce overuse among older adults in three domains:

1. Unnecessary PSA screening in men over 76 years old (PSA); [5,6]
2. Testing women over 65 years old for bacteriuria when no indication is present (UA/UC); [5,7]
3. Overuse of medication to treat diabetes mellitus in adults over 75 years old (DM). [5,8]

To address these aims, we are conducting a two-arm, pragmatic, cluster-randomized controlled trial in an integrated health care system. In this paper, we describe the study design.

2. Methods

2.1. Overview

BEAGLE is a pragmatic, cluster-randomized controlled trial wherein clinics are equally randomized to either receive a set of CDS interventions through their EHR system plus a one-time clinician education module, or simply receive a one-time clinician education module. The intervention period is 18 months followed by a 6-month monitoring phase to monitor for post intervention effects; clinicians completed their education module immediately prior to the start of the intervention period. (Fig. 1).

There are three co-primary outcomes: overtesting for PSA and UA/UC, and overtreatment for DM in older adults, which are assessed with EHR data. We hypothesize that the CDS interventions will reduce the prevalence of these three overuse outcomes. We will also evaluate physician attitudes towards this intervention prior to its start and at the end of the study period.

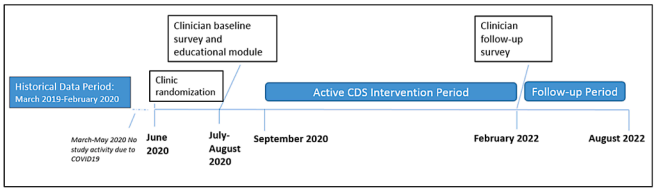

Fig. 1. Study activities.

Since this trial evaluates minimal-risk interventions encouraging clinicians to follow guideline concordant care and outcomes are collected as part of routine care delivery, the Northwestern University Institutional Review Board approved this study with a waiver of informed consent for enrolling clinicians and use of EHR data. A Data and Safety Monitoring Board (DSMB) was convened and approved the study protocol and Data Safety and Monitoring Plan (DSMP) prior to trial initiation. The trial is registered on clinicaltrials.gov [NCT04289753].

2.2. Study setting and eligibility

The study is set in the primary care clinics affiliated with Northwestern Medicine, a large health system in the greater Chicago area in Illinois. The system uses a common EHR, Epic (Epic Systems Corp., Verona, WI), and data are copied nightly to an enterprise data warehouse (EDW). Northwestern Medicine has a system-wide primary care working group with regional representation that approved the study, and medical leaders from the different regions endorsed it.

All primary care internal medicine, family medicine, and geriatrics outpatient clinics affiliated with this system at the time of randomization were eligible for inclusion. Clinics were excluded if they employed only resident physicians or had stopped seeing patients prior to the randomization date.

For study purposes, clinicians were attributed to one participating clinic based on where the plurality of their primary care outpatient clinical activities occurred between March 1, 2019 and February 29, 2020. Clinicians with encounters within at least one randomized clinic were included. We excluded clinicians who were (1) resident physicians, (2) participants in an earlier pilot study of these interventions [18], or (3) study investigators. All eligible clinicians (physicians, advanced nurse practitioners, and physician assistants) were enrolled in the trial with a waiver of informed consent, and their encounters with relevant patients were included in quality measure denominators.

2.3. Study interventions

2.3.1. Clinician education module

All clinicians attributed to participating clinics, regardless of intervention assignment, were emailed a link to an online, interactive educational module covering Choosing Wisely recommendations. The 15–20 min module had three sections related to the study's co-primary outcomes. The module was hosted via Research Electronic Data Capture (REDCap) [19,20], and we tracked module opening and completion.

2.3.2. Clinical decision support interventions

We used a mixed-methods approach to inform our intervention, including both interviews and a feasibility pilot. We piloted EHR CDS interventions designed to reduce overtesting and treatment of older adults. Fourteen primary care physicians consented to participate in a six-month pilot study (January–June 2019) and received CDS alerts for these topics. Prior to designing alerts, we interviewed primary care physicians to understand underlying psychological drivers of overtesting and overtreating older adults. Most physician participants reported knowledge of Choosing Wisely recommendations and could state reasons for not ordering these tests or for de-escalating treatment. However, there appeared to be an underestimation of the potential for harms from unnecessary testing or treatment and an over-valuing of unestablished potential benefits. Additional factors driving clinician behavior included deference to patient preference and yielding to clinical inertia. [21] Thus, we designed CDS interventions to increase the salience of potential harms. We also included content to convey social norms and promote social accountability to conform to the clinical recommendations to not overtest or overtreat. We found the alerts to be clinically accurate, feasible to implement within the EHR, and acceptable to clinicians. Findings from the pilot study provided preliminary

evidence of positive intervention effects. [18]

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.

2.4. Randomization

Study randomization occurred at the clinic level; all 60, eligible primary care clinics seeing adult patients were randomized. We implemented a modified constrained randomization scheme designed to promote balance between study arms on eight clinic-level variables at baseline based on their possible influence on the three co-primary outcomes. The overarching rationale of this approach is to reduce potential bias and promote comparable arms at the analytic unit (clinician/patient encounter) level:

1. Health system region – The clinics are spread around the Chicago area. We categorized each practice into one of four geographic areas (categorical).
2. Number of clinicians – To account for practice size, we included the 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 by a participating clinician (continuous).
4. Annual PSA overtesting rate (percentage).
5. Number of women eligible for unnecessary UA/UC testing (continuous).
6. Annual UA/UC overtesting rate (percentage).
7. Number of patients eligible for DM overtreatment (continuous).
8. Annual DM overtreatment rate (percentage).

These clinic-level variables were constructed using data abstracted from the Epic EHR for the 12 months prior to March 2020 with the exception of six practices that only contributed nine months of data because they had recently joined Northwestern Medicine.

The constrained randomization algorithm we implemented mimics previously presented methodology and was implemented in three steps. [22,23] First, 100,000 candidate 1:1 randomization schemes were generated for the eligible practices, where each clinic was randomized to exactly one intervention strategy. Second, these candidate randomization schemes were then evaluated on imbalance across the eight pre-specified variables. Third, the final randomization scheme was selected at random from the candidate schemes that were 'adequately balanced.'

For a candidate randomization scheme to be considered 'adequately balanced,' the measure of imbalance for each variable 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; [23,24] 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. The study biostatistician generated the final randomization in R version 3.6.4 (R Core Team, 2020). [25] Randomization occurred in June 2020, and CDS interventions were implemented in Epic and the active intervention period began in September 2020. (Fig. 1).

Table 1
Rationale and description of CDS interventions.

Potential driver of clinician testing/ treatment decisions	Behavioral economics/social psychology principle	Topic	Language from CDS interventions presented to clinicians
Underestimation of harm from downstream testing and treatment in a population without established benefits	Make potential downstream harms more available	PSA	"Screening with PSA can lead to harms from diagnostic and treatment procedures."
		UA/ UC	"Warning: Unindicated testing leads to false positives, unnecessary antibiotic treatment, and adverse reactions."
		DM	"Using insulin or sulfonylureas to achieve HbA1c < 7% in most older adults is associated with serious harm, including higher mortality."
Overweighing the risks of not performing the test or treatment (e.g., missing cancer diagnosis)	Provide reassurance that supports the current clinical consensus not to perform the clinical action	PSA	"No guideline recommends routine PSA screening for men over age 75."
		UA/ UC	"Consider not ordering UTI testing initially for non-specific symptoms."
		DM	"Reasonable glycemic targets are…"
Inadequate time within a clinical encounter	Assist clinician by making the preferred decision (cancel order) easy and fast	PSA, UA/ UC	Alert functionality presents option to "Exit and cancel order"
		DM	Alert response options include: "I am reducing treatment now" and "Treatment already reduced."
Inertia/resistance to change (e.g., not wanting to discuss changing screening schedule with patient)	Present social normative information that makes it feel more normal to not perform the action	PSA	"Most NM PCPs use PSA rarely or not at all in men over 75 who have not already been diagnosed with prostate cancer."
	Create sense of accountability to peers; draw on reputational concerns.	PSA, UA/ UC	"If you want to proceed, please enter a justification. If you do not respond, a note indicating 'No justification given for [clinical action] will be added to the chart."

Abbreviations: CDS: clinical decision support; PSA: prostate-specific antigen; UA/UC: urine studies; DM: type 2 diabetes; UTI: urinary tract infection; NM: Northwestern Medicine; PCP: primary care physician.

2.5. Data collection

All intervention and outcome data come directly from EHR data documented as part of routine care delivery. Each CDS intervention has a unique identifier that tracks when it is triggered and a clinician's response to each alert including their discrete clicks (e.g., to proceed or cancel order) and free text entries within the CDS. Queries of the EDW for intervention delivery and study outcome data were validated against clinician chart review.

The interventions are delivered automatically to the clinician user in the EHR whenever the triggering criteria are met within a clinical encounter. We expect that, based on the type of intervention (automated/computerized), the intervention will be delivered with 100% fidelity.

The major outcomes of this study are assessed using electronic queries of data in the EDW copied from the EHR. There is no difference in the assessment methodology applied to the intervention arm and the education-only arm, and human interpretation is not part of the outcome assessment.

2.5.1. Clinician survey

To better understand clinicians' baseline attitudes, all clinicians attributed to clinics in both study arms were recruited via email to complete a survey during July and August 2020, prior to activating CDS interventions. The survey (included in the supplemental material) was administered through REDCap and took 15–20 min to complete. [19] Clinicians provided informed consent prior to accessing the survey. The survey included items about clinician knowledge of guidelines, case studies, perceptions of their clinical treatment decisions for their older adult patients, perceptions of their peer clinicians' behavior, and demographic items. Clinicians who completed the survey received $200 in compensation. A follow-up survey is planned following the end of the 18-month intervention period to enable within-clinician pre-post comparisons.

2.6. Study measures

2.6.1. Co-primary outcomes

Table 2 details the three co-primary outcomes for this trial.

To validate outcome measures, we performed EDW queries, and then a physician performed manual chart review on a random sample to assess accuracy. We sampled 20 records that did and did not meet each measurement criteria (e.g. sex, age, encounter) that comprised the specifications for the numerator, denominator and exclusions for each of the three outcomes. When chart review identified clinical content criteria that had not been included in the measure criteria (i.e., a potentially valid diagnosis that should be included as an exclusion criteria), two physicians discussed and reached consensus about whether to modify the criteria for the outcome measure. In cases where the query results did not match the chart, the physician reviewer and the analyst worked together to identify the source of the discrepancy, and the analyst modified the query logic when appropriate. We repeated the automated queries and performed additional physician review to determine that identified errors had been corrected in all 20 records.

To investigate durability of effects, we will additionally evaluate the three co-primary outcomes in the 6 months immediately post-intervention. Detailed specifications for all outcomes are included in the supplemental material. Other outcomes were pre-specified and listed within the clinicaltrials.gov record prior to trial initiation.

2.6.2. Clinician attitude measures

Secondary outcomes include clinician attitudes and beliefs about testing and treatment for older adults and their opinions on this study's CDS interventions among all clinicians who completed study surveys at the end of follow-up.

Table 2
Description of primary outcome measures in older patients for BEAGLE Trial.

Measure	Eligibility criteria	Outcome description
PSA over-screening (PSA)	Men aged 76 years and older at the start of the measurement period with: (1) at least one visit during the measurement period with an included clinician; (2) no diagnosis or procedure code suggesting a history of prostate cancer; (3) no prescription for any androgenic agents	The presence of a PSA laboratory result in the EHR during the measurement period
Urine testing for non-specific reasons (UA/UC)	Women 65 years or older at the start of the measurement period where a study clinician obtained a urinalysis (UA) and/or urine culture (UC) in the interval 24 h before to 48 h after a face-to-face ambulatory care visit[a]	The absence of a diagnostic code for a specific genitourinary sign, symptom or other potentially relevant indication
Diabetes overtreatment (DM)	Adults aged 75 years and older with a diagnosis of diabetes mellitus (DM)	Patients with (1) their most recent hemoglobin A1C during the measurement period of less than 7.0[b] and (2) insulin, a sulfonylurea or meglitinide on their active medication list at the end of the measurement period.

[a] If more than one qualifying visit, only the first testing episode in a measurement period is included.
[b] If no hemoglobin A1C was obtained during the measurement period, carry forward hemoglobin A1C from the year prior to the measurement period.

2.6.3. Safety outcomes

Safety outcomes are defined and included in the trial's DSMP. There are two safety measures for the UA/UC quality measure: (1) Urinary tract infection (UTI) requiring hospital care following CDS exposure, and (2) UTI requiring hospital care rate among women 65 years and older following an office visit (comparison between study arms). There are three safety measures related to the DM quality measure: (1) Hyperglycemia requiring hospital care following CDS exposure, (2) Hyperglycemia requiring hospital care rate among previously tightly controlled patients (a recent HbA1C < 7) (comparison between study arms), and (3) Poor diabetes control among previously tightly controlled patients (comparison between study arms).

2.7. Statistical analysis plan

The primary outcomes for BEAGLE are measures of clinician overtreatment in three areas: PSA, UA/UC, and DM. Individual patient-level data will be used as the unit of analysis; the unit of randomization is the clinic. We will analyze each measure separately as different patient populations are eligible for each outcome. Analyses will focus on evaluating changes across arms in these outcomes from baseline to 18 months. Secondary analyses will explore within practice mean changes in each outcome.

It is plausible that during the study period, we might observe a decreasing trend over time in overtreatment of older adults even in the contemporary control arm. Therefore, we will investigate trends during the period immediately preceding study start to evaluate historical controls, and explicitly model time in our primary outcome models. Specifically, we will test the impact of the intervention using a piecewise hierarchical mixed effects logistic regression model with a knot at 0 month (start of intervention) that will include random effects for clinics and clinicians:

$$logit\{Pr(Y_{ijkt}=1)\}=\beta_0+\beta_{0j}+\beta_{0k}+\beta_1A_k+\beta_2t+\beta_3A_kt+\beta_4A_k\times I(t\geq 0)+\beta_5t\times I(t\geq 0)+\beta_6A_kt\times I(t\geq 0)+\overrightarrow{\beta_7W_{ijkt}}+\epsilon_{ijkt}. \quad (1)$$

Here, Y_{ijkt} represents the outcome (binary) for the i^{th} person seen by the j^{th} clinician at the k^{th} clinic in month t, and A_k represents the assigned study arm for clinic k where the j^{th} clinician was attributed. [9] We will assume that all random effects are normally distributed and independent. We have included both clinician-level and clinic-level random intercepts. We will use separate parametric bootstraps to test at the $\alpha = 0.05$ level whether these random effects add information to the model; if not, we will remove them for parsimony. The estimate of interest is β_6, which represents the difference in log-odds of overtreatment between study arms during the study period. Analyses will be adjusted for monthly values of the characteristics that were included in the constrained randomization algorithm, represented by W_{ijkt} in Eq. 1. Standard errors will be estimated using a nonparametric bootstrap with 1000 resamplings.

Due to the interruption to primary practice caused by COVID-19, our historical data period will exclude data from March 01, 2020 through August 31, 2020 (Fig. 1). Our outcome models will include data from a 12 month historical control period from March 1, 2019 through February 29, 2020, and then during the 18 month intervention period, September 1, 2020 through February 28, 2022.

All primary analyses will be based on the intention-to-treat principle (i.e., all randomized clinics and patients from all clinicians included, according to the arm to which they were randomized). We plan to additionally conduct a set of sensitivity analyses to estimate the per-protocol effect. These analyses will account for loss to follow-up. Additional sensitivity analyses will assume no trend in time before or during the study time (removing the linear time terms from Eq. 1); in those analyses the equivalent of β_4 will be the estimate of interest. We will further replicate all three primary analyses separately by patient race/ethnicity, and test for interaction by sex for the DM analysis. Analyses will examine each measure separately, and will assume a two-sided 5% level of significance implying that each primary outcome estimate will be considered significant at $p < 0.017$ (Bonferroni correction for multiple hypothesis tests). [26,27]

Regular quality checks will be performed to ensure that there are no issues with the EHR data, such as systematically missing data due to updates to the EHR system that interfere with the data extraction workflow. All analyses will be conducted using SAS v9.4 (SAS Institute, Casey, NC) or R v4.1.1 (R Core Team) software. [25,28]

2.8. Sample size and power

We structured our power calculations 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. [29] Historic data 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 0.017), 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 of 20–30% for PSA over-screening and UA/UC overtesting, and 15–25% for DM overtreatment, and 9480, 15,000, and 6600 patients eligible for each outcome analysis respectively. These analyses should be conservative, as our assumed sample size is from an historic 12-month period, and our follow-up period will be 18 months. We did not assume any clinic drop-out, as all data will be available via the EHR. With these assumptions, we anticipated that we have adequate power to detect an absolute difference of roughly 4% for each of the measures. For the PSA and UA/UC outcomes, this corresponds to a relative risk between 0.80 and 0.87; for the DM outcome, it corresponds to a relative risk between 0.73 and 0.84.

3. Results

There were 60 eligible clinics enrolled into the study. Of the 100,000 candidate randomization schemes generated, 42 satisfied balancing criteria, and one was randomly selected for implementation. (Fig. 2) Table 3 contains summary statistics for the clinic characteristics overall and stratified by study arm. Clinics had, on average, 6.3 clinicians attributed to them. In the year prior to randomization, the rates of PSA overscreening, UA/UC overtesting, and DM overtreatment were 24.9%, 23.9%, and 16.8%, respectively. Participating clinicians were 63% female, 86% physicians, and the majority (65%) had an Internal Medicine specialty. (Table 4).

We achieved adequate balance for seven of the eight pre-specified variables used in the constrained randomization algorithm. The number of clinics in the four geographic regions is identical between arms, as is the average number of clinicians (6). The annual UA/UC and DM measures – specifically, the number of eligible patients seen in clinic in the year prior to randomization (denominator), and the percentage of those patients who were overtested or overtreated – also achieved adequate balance. However, due to a coding issue, we inadvertently applied balancing criteria for the PSA measure on the number of eligible patients (denominator) and the number of PSA orders (numerator) rather than the percentage of older men who received PSA screening – an absolute measure instead of a relative measure that accounts for the size of the eligible patient population at each clinic. Unfortunately, this resulted in less than what we had a priori deemed 'adequate' balance on the intended characteristic – the percentage of older men who received PSA screening (31% versus 19%). While we acknowledge this oversight, we suspect that this will not bias our intervention effect estimates in analyses for a few reasons: (1) the variable we used in the algorithm is correlated with the desired variable (percentage of older men who

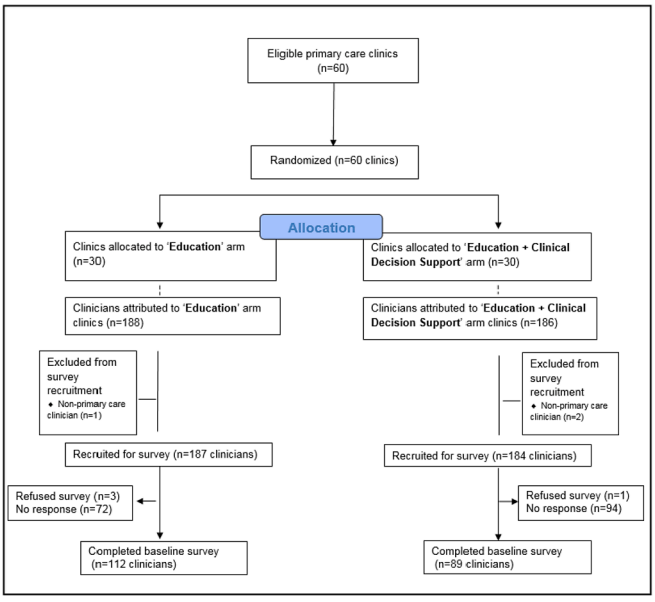

Fig. 2. Clinic enrollment and clinician attribution and baseline survey participation diagram in the BEAGLE study as of August 2020.

Table 3
Clinic characteristics and baseline ordering of tests and diabetes overtreatment prior to BEAGLE randomization: March 2019 – February 2020.

	Overall *N* = 60 clinics *n* = 377 clinicians	**Intervention *N* = 30 clinics *n* = 185 clinicians**	**Education only N = 30 clinics *n* = 192 clinicians**
	Mean (SD)	Mean (SD)	Mean (SD)
No. of Clinicians per Clinic	6.3 (6.1)	6.2 (5.7)	6.4 (6.5)
No. of patients eligible for unnecessary PSA screening per clinician	155.4 (167.4)	149.3 (182.1)	161.7 (153.8)
No. of patients with PSA results[a]	37.1 (45.3)	38.0 (47.0)	36.0 (44.2)
PSA screening in older men (by PCPs), %[a]	24.9 (19.1)	30.9 (23.3)	18.7 (10.8)
No. of patients eligible for unnecessary UA/UC testing	239.0 (304.5)	252.3 (338.3)	225.6 (271.7)
No. of patient with urine tested for non-specific reasons	68.0 (115.3)	80.6 (141.7)	55.6 (81.5)
Urine testing for non-specific reasons (UA/UC),%	23.9 (13.7)	23.3 (13.4)	24.6 (14.1)
No. of patients eligible for diabetes overtreatment	117.5 (127.6)	111.8 (134.1)	123.5 (122.6)
No. of older adults with overtreated diabetes	21.9 (26.2)	22.4 (30.3)	21.3 (21.6)
Diabetes overtreatment in older adults,%	16.8 (9.7)	16.8 (10.3)	16.9 (9.2)
Health system region, N (%)			
Region1	14 (23.3)	7 (23.3)	7 (23.3)
Region2	8 (13.3)	4 (13.3)	4 (13.3)
Region3	6 (10.0)	3 (10.0)	3 (10.0)
Region4	32 (53.3)	16 (53.3)	16 (53.3)

Abbreviations: BEAGLE: Behavioral Economic Applications to Geriatrics Leveraging Electronic Health Records study; SD: standard deviation; PSA: prostate-specific antigen; PCP: primary care provider; UA/UC: urine studies.

[a] Unmatched PSA rate is explained by error in constrained randomization using PSA numerator count as opposed to rate variable.

Table 4
Baseline clinician characteristics: BEAGLE Study, July – August 2020.

	Overall	**Intervention**	**Education only**
Number of Clinics	60	30	30
Number of Clinicians	371	184	187
Clinician specialty, %			
Family Medicine	32.6	34.8	30.5
Geriatrics	2.2	1.6	2.7
Internal Medicine	65.2	63.6	66.8
Clinician type, %			
Physician	85.7	88.6	82.9
Physician Assistant	6.7	3.8	9.6
Advanced Practice Nurse	7.5	7.6	7.5
Female, mean (SD)	232 (62.5)	117 (63.6)	115 (61.5)

Abbreviations: BEAGLE: Behavioral Economic Applications to Geriatrics Leveraging Electronic Health Records study.

received PSA screening), (2) the final model to evaluate the effectiveness of the intervention will adjust for desired measure of interest at baseline along with all variables pre-specified for the constrained randomization process. The constrained randomization process may be viewed as a protection against large chance imbalances that may occur at random in any of these variables. Through the process we have outlined here, we have constrained this full 'randomization space' of 100,000 possible schemes to just 42 (or to the optimal 0.042% according to these pre-specified thresholds). [30]

4. Discussion

Wide variation exists in the care provided to older adults, even when they are treated within the same system. [31] Here, we used a modified constrained randomization process to allocate 60 clinics to either receive clinician education alone or clinician education plus CDS interventions informed by behavioral economics and social psychology aimed at improving care for older adults. The baseline overuse rates were 25% for the PSA measure, 23% for the UA/UC measures, and 17% for the DM overtreatment measure. These baseline rates highlight that opportunities exist to reduce overtesting and treatment among older adults.

This large pragmatic randomized controlled trial has the potential to provide evidence of the effects of using CDS informed by behavioral science to curb overuse in ambulatory care of older adults. Trial results will reference this manuscript to summarize trial design. Data collection and intervention delivery will continue through 2022 and final analyses and primary results will be disseminated after trial conclusion.

This trial of previously pilot-tested CDS interventions has several strengths, including sufficient power to detect clinically-meaningful differences, novel constrained randomization scheme, and highly pragmatic design with simple implementation. There are several aspects to this implementation we particularly call attention to for other researchers. We chose to do the study in a system that has multiple features conducive to this type of work: 1) strong working relationships between operational clinical informatics and quality improvement leaders and researchers, 2) governance systems that allowed for approval of the study across the health system's primary care practices with a waiver of individual informed consent, and 3) primary care leadership that recognizes and is supportive of the concept of a learning health system and using routine care delivery to generate new knowledge. This enabled us to test these approaches having them appear to clinicians in a fashion similar to other routine CDS and to include all of the candidate primary care practices.

We acknowledge as a limitation that this study is conducted within a single large health system on one EHR platform so findings may not be generalizable to other settings.

Results of this study will contribute to current understanding about the power of incorporating behavioral economics and social psychology principles into CDS tools in modern EHR systems to improve quality of care for older adults.

Funding source

Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under Award Number R33AG057383. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. REDCap is supported at Feinberg School of Medicine by the Northwestern University Clinical and Translational Science (NUCATS) Institute. Research reported in this publication was supported, in part, by the National Institutes of Health's National Center for Advancing Translational Sciences, Grant Number UL1TR001422. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health in publications relating to this project. The NIH had no role in the study design, in the collection, analysis and interpretation of data, in the writing of the report, and in the decision to submit the article for publication.

Declaration of Competing Interest

Dr. Persell, Dr. Petito, Ms. Lee and Mr. Peprah receive research support from Omron Healthcare. Dr. Persell, Ms. Lee and Mr. Peprah previously received research support from Pfizer, Inc. Dr. Persell is a consultant for QuiO Technologies and the RAND Corporation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cct.2021.106649.

References

[1] D.M. Levine, J.A. Linder, B.E. Landon, The quality of outpatient care delivered to adults in the United States, 2002 to 2013, JAMA Intern. Med. 176 (12) (2016) 1778–1790.
[2] P. Rotjanapan, D. Dosa, K.S. Thomas, Potentially inappropriate treatment of urinary tract infections in two Rhode Island nursing homes, Arch. Intern. Med. 171 (5) (2011) 438–443.
[3] D.E. Bonds, et al., The association between symptomatic, severe hypoglycaemia and mortality in type 2 diabetes: retrospective epidemiological analysis of the ACCORD study, BMJ 340 (2010), b4909.
[4] R. Besdine, et al., Caring for older Americans: the future of geriatric medicine, J. Am. Geriatr. Soc. 53 (6 Suppl) (2005) S245–S256.
[5] ABIM, Choosing Wisely, An Initiative of the ABIM Foundation [07/01/2021]; Available from, http://www.choosingwisely.org/wp-content/uploads/2015/02/AGS-Choosing-Wisely-List.pdf, 2015.
[6] R. Etzioni, et al., Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends, J. Natl. Cancer Inst. 94 (13) (2002) 981–990.
[7] L.E. Nicolle, et al., Infectious Diseases Society of America guidelines for the diagnosis and treatment of asymptomatic bacteriuria in adults, Clin. Infect. Dis. 40 (5) (2005) 643–654.
[8] K.J. Lipska, et al., Potential overtreatment of diabetes mellitus in older adults with tight glycemic control, JAMA Intern. Med. 175 (3) (2015) 356–362.
[9] D. Meeker, et al., Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: a randomized clinical trial, JAMA 315 (6) (2016) 562–570.
[10] S.D. Persell, et al., Behavioral interventions to reduce inappropriate antibiotic prescribing: a randomized pilot trial, BMC Infect. Dis. 16 (2016) 373.
[11] J.A. Linder, et al., Effects of behavioral interventions on inappropriate antibiotic prescribing in primary care 12 months after stopping interventions, JAMA 318 (14) (2017) 1391–1392.
[12] B.J. Wright, et al., Low-cost behavioral nudges increase Medicaid take-up among eligible residents of Oregon, Health Aff. 36 (5) (2017) 838–845.
[13] H.M. Belli, et al., Implementation of a behavioral economics electronic health record (BE-EHR) module to reduce overtreatment of diabetes in older adults, J. Gen. Intern. Med. 35 (11) (2020) 3254–3261.
[14] D. Kahneman, et al., Prospect theory: an analysis of decision making under, Risk. 43 (1977).
[15] A. Tversky, D. Kahneman, Availability: a heuristic for judging frequency and probability, Cogn. Psychol. 5 (2) (1973) 207–232.
[16] R.B. Cialdini, Influence: science and practice, Prentice Hall. (2009) 260.
[17] J.S. Lerner, P.E. Tetlock, Accounting for the effects of accountability, Psychol. Bull. 125 (2) (1999) 255–275.
[18] T.A. Rowe, et al., Behavioral economics informed decision support to improve choosing wisely adherence when ordering tests and treatments for geriatric patients: a pilot study, J. Gen. Intern. Med. 35 (Suppl. 1) (2020) S44.
[19] P.A. Harris, et al., Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support, J. Biomed. Inform. 42 (2) (2009) 377–381.
[20] P.A. Harris, et al., The REDCap consortium: building an international community of software platform partners, J. Biomed. Inform. 95 (2019), 103208.
[21] T.A. Rowe, et al., Examining primary care physician rationale for not following geriatric choosing wisely recommendations, BMC Fam. Pract. 22 (1) (2021) 95.
[22] P.J. Nietert, et al., An application of a modified constrained randomization process to a practice-based cluster randomized trial to improve colorectal cancer screening, Contemp. Clin. Trials 30 (2) (2009) 129–132.
[23] J.D. Ciolino, et al., Design of healthy hearts in the heartland (H3): a practice-randomized, comparative effectiveness study, Contemp. Clin. Trials 71 (2018) 47–54.
[24] W. Zhao, M.D. Hill, Y. Palesch, Minimal sufficient balance-a new strategy to balance baseline covariates and preserve randomness of treatment allocation, Stat. Methods Med. Res. 24 (6) (2015) 989–1002.
[25] R Foundation for Statistical Computing, Vienna, Austria, The R Project for Statistical Computing, Available from, https://www.R-project.org/.
[26] O.J. Dunn, Multiple comparisons among means, J. Am. Stat. Assoc. 56 (293) (1961) 52–64.
[27] T.J. VanderWeele, M.B. Mathur, Some desirable properties of the Bonferroni correction: is the Bonferroni correction really so bad? Am. J. Epidemiol. 188 (3) (2019) 617–618.
[28] SAS Institute, Casey, NC, SAS Software [06/01/2020]; Available from, sas.com/en_us.
[29] NCSS, LLC, Kaysville, Utah, USA. *PASS 15 Power Analysis and Sample Size Software* [06/01/2020]; Available from, ncss.com/software/pass, *2017*.
[30] M. Organ, et al., Evaluating performance of covariate-constrained randomization (CCR) techniques under misspecification of cluster-level variables in cluster-randomized trials, Contemp Clin Trials Commun 22 (2021), 100754.
[31] T.A. Rowe, et al., Clinician-level variation in three measures representing overuse based on the American Geriatrics Society choosing wisely statement, J. Gen. Intern. Med. 35 (6) (2020) 1797–1802.