Original Investigation

Effect of Behavioral Interventions on Inappropriate Antibiotic Prescribing Among Primary Care Practices
A Randomized Clinical Trial

Daniella Meeker, PhD; Jeffrey A. Linder, MD, MPH; Craig R. Fox, PhD; Mark W. Friedberg, MD, MPP; Stephen D. Persell, MD, MPH; Noah J. Goldstein, PhD; Tara K. Knight, PhD; Joel W. Hay, PhD; Jason N. Doctor, PhD

IMPORTANCE Interventions based on behavioral science might reduce inappropriate antibiotic prescribing.

OBJECTIVE To assess effects of behavioral interventions and rates of inappropriate (not guideline-concordant) antibiotic prescribing during ambulatory visits for acute respiratory tract infections.

DESIGN, SETTING, AND PARTICIPANTS Cluster randomized clinical trial conducted among 47 primary care practices in Boston and Los Angeles. Participants were 248 enrolled clinicians randomized to receive 0, 1, 2, or 3 interventions for 18 months. All clinicians received education on antibiotic prescribing guidelines on enrollment. Interventions began between November 1, 2011, and October 1, 2012. Follow-up for the latest-starting sites ended on April 1, 2014. Adult patients with comorbidities and concomitant infections were excluded.

INTERVENTIONS Three behavioral interventions, implemented alone or in combination: *suggested alternatives* presented electronic order sets suggesting nonantibiotic treatments; *accountable justification* prompted clinicians to enter free-text justifications for prescribing antibiotics into patients' electronic health records; *peer comparison* sent emails to clinicians that compared their antibiotic prescribing rates with those of "top performers" (those with the lowest inappropriate prescribing rates).

MAIN OUTCOMES AND MEASURES Antibiotic prescribing rates for visits with antibiotic-inappropriate diagnoses (nonspecific upper respiratory tract infections, acute bronchitis, and influenza) from 18 months preintervention to 18 months afterward, adjusting each intervention's effects for co-occurring interventions and preintervention trends, with random effects for practices and clinicians.

RESULTS There were 14 753 visits (mean patient age, 47 years; 69% women) for antibiotic-inappropriate acute respiratory tract infections during the baseline period and 16 959 visits (mean patient age, 48 years; 67% women) during the intervention period. Mean antibiotic prescribing rates decreased from 24.1% at intervention start to 13.1% at intervention month 18 (absolute difference, −11.0%) for control practices; from 22.1% to 6.1% (absolute difference, −16.0%) for suggested alternatives (difference in differences, −5.0% [95% CI, −7.8% to 0.1%]; *P* = .66 for differences in trajectories); from 23.2% to 5.2% (absolute difference, −18.1%) for accountable justification (difference in differences, −7.0% [95% CI, −9.1% to −2.9%]; *P* < .001); and from 19.9% to 3.7% (absolute difference, −16.3%) for peer comparison (difference in differences, −5.2% [95% CI, −6.9% to −1.6%]; *P* < .001). There were no statistically significant interactions (neither synergy nor interference) between interventions.

CONCLUSIONS AND RELEVANCE Among primary care practices, the use of accountable justification and peer comparison as behavioral interventions resulted in lower rates of inappropriate antibiotic prescribing for acute respiratory tract infections.

TRIAL REGISTRATION clinicaltrials.gov Identifier: NCT01454947

← Editorial page 558

+ Supplemental content at jama.com

+ CME Quiz at jamanetworkcme.com and CME Questions page 608

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Jason N. Doctor, PhD, Leonard D. Schaeffer Center for Health Policy and Economics, School of Pharmacy, University of Southern California, Verna & Peter Dauterive Hall (VPD), 635 Downey Way, Los Angeles, CA 90089-3333 (jdoctor@usc.edu).

Overuse of antibiotics exposes patients to unnecessary risk of adverse drug events, increases health care costs, and increases the prevalence of antibiotic-resistant bacteria.[1,2] Most antibiotics prescribed in the United States are for acute respiratory tract infections, and roughly half of these prescriptions are intended to treat diagnoses for which antibiotics have no benefit.[3,4] Despite published clinical guidelines and decades of efforts to change prescribing patterns, antibiotic overuse persists.[1,2,5,6] Interventions such as physician and patient education, computerized clinical decision support, and financial incentives have historically produced modest reductions in antibiotic prescription rates for targeted acute respiratory tract infections.[7,8]

Changing clinician decision making has been challenging in acute respiratory visits and other care domains. For example, pay-for-performance has yielded mixed results,[9,10] and traditional alerts and reminders, which can contribute to information overload, are often disruptive and ignored.[11] Two recent studies favored an intervention that did not disrupt workflow and that appealed to clinicians' pride in their own performance.[12,13] With this in mind, there is increasing interest in use of behavioral science, including psychology and behavioral economics, to affect policy.[14] Researchers are beginning to apply models from these disciplines to identify new social and cognitive devices to gently nudge clinician decision making while preserving freedom of choice.[15] Such approaches are well matched to the goal of encouraging uptake of effective evidence-based treatments in health care, with appropriate antibiotic prescribing being one example.

We applied insights from behavioral science to design 3 interventions to reduce the rate of unnecessary antibiotic prescribing during ambulatory visits for acute respiratory tract infections in a multisite cluster randomized trial.

Methods

Study Design
Details of the study design, randomization scheme, and interventions are summarized below and presented in detail in the original protocol (available in Supplement 1). The institutional review board of each participating institution approved all study procedures and waived informed consent for patients.

Settings
We recruited 49 primary care practices from 3 health systems using 3 different electronic health records (EHRs) in 2 geographically distinct regions: Massachusetts (Partners Health-Care: 22 practices affiliated with Brigham and Women's Hospital or Massachusetts General Hospital) and Southern California (AltaMed Medical Group, 22 practices; The Children's Clinic, 5 practices [the latter also sees a high volume of adult patients]). Participating practices served different patient populations. The majority of patients at the Southern California practices were Hispanic, with a high proportion living at or below 200% of the federal poverty level, whereas the patients served by the Massachusetts practices were predominantly white/non-Hispanic, with a wider income range. Race/ethnicity was assessed as part of the standard collection of demographic data from EHRs, and harmonized using Observational Medical Outcomes Partnership standards for race/ethnicity reporting.[16]

Practices were excluded prior to randomization if none of their clinicians had at least 5 antibiotic-inappropriate acute respiratory tract infection visits annually. Two of the practices did not meet inclusion criteria, leaving 47 practices for randomization. All sites required clinicians to prescribe medications through their EHRs.

Enrollment
Clinicians were recruited via email and enrolled through an online education module covering acute respiratory tract infection diagnosis and treatment guidelines, information about the interventions, and any expected changes to their EHRs (details in the original protocol, available in Supplement 1); clinicians provided electronic informed consent at module completion. Some organizations allowed payment to individual clinicians for their participation; others instead requested payment to practices. Payment for study participation was $1200 per clinician, regardless of intervention assignment or antibiotic prescribing rates.

Outcome Measures
The primary study outcome was the antibiotic prescribing rate for antibiotic-inappropriate acute respiratory tract infection visits and no concomitant reason for antibiotic prescribing. Antibiotic-inappropriate diagnoses included nonspecific upper respiratory tract infections, acute bronchitis, and influenza (*International Classification of Diseases, Ninth Revision* [*ICD-9*] codes 460, 464, 464.0, 464.00, 464.1, 464.10, 464.2, 464.20, 464.4, 464.50, 465, 465.0, 465.8, 465.9, 466, 466.0, 466.1, 466.11, 466.19, 487, 487.1, 487.8, 490). We excluded visits with diagnosis codes for acute pharyngitis or acute rhinosinusitis because guidelines permit antibiotic prescription when certain criteria are met, and we lacked data necessary to identify this antibiotic-appropriate subset.[17,18]

A visit for an antibiotic-inappropriate acute respiratory tract infection was eligible for outcome inclusion if (1) the patient was 18 years or older, (2) the clinician and practice were enrolled in the study, (3) the visit occurred during the 18-month baseline or 18-month intervention period, and (4) the patient had no visit for acute respiratory tract infection within the prior 30 days. Visits were excluded when patients had medical comorbidities that were acute respiratory tract infection guideline exclusions (eg, chronic lung disease; for full list of excluded diagnoses, see original protocol [Appendix E: Code Set Definitions] in Supplement 1) or patients had concomitant visit diagnoses indicating presence of other, potentially antibiotic-appropriate, infections (eg, cellulitis, acute sinusitis).

Data Abstraction
Each participating site created an extract from its EHR or billing records of the data elements necessary to compute the outcome measures for all patients with acute respiratory tract infections. These records were transferred to the University of Southern Cali-

fornia, where study staff checked data quality and transformed the data into a standard model (Observational Medical Outcomes Partnership Common Data Model, version 3).[16]

Interventions

Suggested alternatives was an EHR-based intervention most closely resembling traditional clinical decision supports and order sets. Diagnoses of acute respiratory tract infection triggered a pop-up screen stating that "Antibiotics are not generally indicated for [this diagnosis]. Please consider the following prescriptions, treatments, and materials to help your patient," followed by a list of alternatives (see original protocol [Appendix F: Example of Suggested Alternatives Order Set] in Supplement 1), each with streamlined order entry options for over-the-counter and prescription medications (eg, decongestants) and letter templates excusing patients from work. The suggested alternatives intervention drew from the behavioral insight that prescribers may infer that a suggested (nonantibiotic) alternative ought to be considered, thus reducing the likelihood that an antibiotic would be prescribed.[19]

Accountable justification was also an EHR-based intervention. An EHR prompt asked each clinician seeking to prescribe an antibiotic to explicitly justify, in a free text response, his or her treatment decision. The prompt also informed clinicians that this written justification would be visible in the patient's medical record as an "antibiotic justification note" and that if no justification was entered, the phrase "no justification given" would appear. Encounters could not be closed without the clinician's acknowledgment of the prompt, but clinicians could cancel the antibiotic order to avoid creating a justification note, if they chose. The accountable justification alert was triggered for both antibiotic-inappropriate diagnoses and potentially antibiotic-appropriate acute respiratory tract infection diagnoses (eg, acute pharyngitis).

The accountable justification intervention was based on prior findings that accountability improves decision making accuracy and that public justification engenders reputational concerns.[20-23] To preserve their reputations, clinicians should be more likely to act in line with injunctive norms[24]—that is, what one "ought to do" as recommended by clinical guidelines.[25]

Peer comparison was an email-based intervention. Clinicians were ranked from highest to lowest inappropriate prescribing rate within each region using EHR data. Clinicians with the lowest inappropriate prescribing rates (the top-performing decile) were told via monthly email they were "Top Performers" (see original protocol [Appendix G: Sample Peer Comparison Email Text] in Supplement 1). The remaining clinicians were told that they were "Not a Top Performer" in an email that included the number and proportion of antibiotic prescriptions they wrote for antibiotic-inappropriate acute respiratory tract infections, compared with the proportion written by top performers.

Peer comparison was distinct from traditional audit-and-feedback interventions in its comparison with top-performing peers instead of average-performing peers and its delivery of positive reinforcement to top performers—a strategy shown elsewhere to sustain performance.[26-28]

Intervention Triggers

The suggested alternatives and accountable justification interventions were triggered by clinician entry of antibiotic orders for antibiotic-inappropriate acute respiratory tract infections and those for which antibiotics were potentially (but not necessarily) appropriate: acute rhinosinusitis and acute pharyngitis. The peer comparison intervention included only antibiotic-inappropriate acute respiratory tract infections. All 3 interventions were suppressed for patients who had comorbidities constituting guideline exclusions.

Randomization

We randomized practices to receive 0, 1, 2, or 3 interventions in a 2 ×2 × 2 factorial design to avoid within-practice contamination between clinicians, blocking on geographic region (in part, to balance study allocation by EHR) to achieve balance for evaluation of main effects.[29] This factorial design enabled investigation of potential interactions between intervention effects, while preserving the ability to estimate the main effects of each intervention individually. Randomization programming was conducted with the R statistical programming package (R Foundation for Statistical Computing; https://www.r-project.org/foundation/).

Implementation

Dates for intervention implementation differed between practice organizations, owing to differences in clinician recruitment procedures and EHR-specific development times for intervention features. Interventions began between November 2011 and October 2012 and lasted for 18 months in each practice. For the latest-starting practices, follow-up ended on April 1, 2014. A study timeline is presented in the eFigure in Supplement 2.

Statistical Analysis

Our primary analysis was based on the approach of Gerber et al[12]: we used a piecewise hierarchical model with a knot at month 0 (the intervention start date) to model trajectories of the log odds of the main outcome for the control group and each intervention group, starting 18 months before each intervention began and ending after 18 months of intervention exposure, and including random effects for practices and clinicians. The estimates of interest were intervention × time interaction terms, which represented changes in prescribing trajectories, relative to contemporaneous controls, that occurred when each intervention began. This model measured the effects of each intervention in comparison with all practices that did not receive the intervention, adjusting for exposure to other interventions and practice- and clinician-level effects to account for time-invariant characteristics (eg, specific EHR product used). We did not adjust for patient characteristics, which were measured after clinicians were exposed to the interventions and could be concomitant with outcome. Instead, we relied on block randomization to equate groups on patient characteristics.

To display findings on the original scale of the data, we generated monthly marginal predictions and confidence intervals from this model corresponding to the control condition and each intervention individually. Confidence intervals for

differences between control and intervention (ie, for intervention effects) were bootstrapped with 1000 replications.

We performed sensitivity analyses to test for interactions between interventions by expanding the main effects model to include interaction terms for each combination of interventions and comparing this fully interacted model to the original main model using a Wald test.[30] We also evaluated interaction terms individually using a similar approach.

With an expected 2252 or more visits per intervention group, a priori calculations indicated 80% power to detect a 7% absolute reduction in antibiotic prescribing (less than the 8.9% median effect in prior efforts to improve antibiotic quality improvement)[8] at the .05 level of significance, assuming a baseline prescribing rate of 50% and intrapractice correlation coefficient of 0.05. These expected visit counts and correlation coefficients were based on preintervention analyses, using data from the Boston and Los Angeles sites.

To investigate the possibility that interventions led to diagnosis shifting (ie, changes in clinicians' diagnostic coding habits), we used the approach of our main models to test whether potentially antibiotic-appropriate acute respiratory tract infection diagnoses (eg, pneumonia, chronic sinusitis) increased as a proportion of all acute respiratory tract infection diagnoses.[31]

In sensitivity analyses, we fit a simple difference-in-differences model estimating changes in the primary outcome associated with each intervention, treating the entire 18-month intervention period as a binary variable, without accounting for prescribing trajectories.

Elements of our analytic approach (specifically, using an 18-month baseline period and piecewise hierarchical modeling technique in the main analyses, performing the interaction effect sensitivity analysis, and performing the simple difference-in-differences model sensitivity analyses) were modified from our original analysis plan based on feedback received during the peer review process.

Under the direction of an independent data and safety monitoring board, we evaluated patient safety. For antibiotic-inappropriate visits in which no antibiotic was prescribed, we assessed return visits within 30 days for the presence of complications potentially attributable to untreated bacterial infections (eTable 1 in Supplement 2). We conducted chart reviews on a 20% random sample of such cases to determine whether prescription of antibiotics at the initial visit would have prevented the complication.

We performed a complete case analysis. There were no missing values of the main outcome, and we did not impute missing covariate values (which were missing in approximately 3% of records). We analyzed data using Stata MP version 12.1 (StataCorp) and considered 2-sided *P* values less than .05 significant.

Results

Sample Characteristics
Of 353 clinicians invited, 248 (70%) agreed to participate (**Figure 1**). On average, enrolled clinicians were 48 years old,

and most were women (**Table 1**). Among patients who had a qualifying visit during the 18-month intervention period, the mean age was 48 years, 33% were men, 87% were white, 32% were Hispanic, and 59% had private insurance (patient characteristics during the baseline period are available in eTable 2 in Supplement 2).

During the study, there were 125 333 visits for any diagnosis of acute respiratory tract infection. Of these, 31 712 visits (14 753 during baseline period, 16 959 during intervention period) met criteria for outcome evaluation.

Antibiotic Prescribing
Mean antibiotic prescribing rates decreased from 24.1% at intervention start to 13.1% at intervention month 18 (absolute difference, −11.0%) for control practices; from 22.1% to 6.1% (absolute difference, −16.0%) for suggested alternatives (difference in differences, −5.0% [95% CI, −7.8% to 0.1%]; *P* = .66 for differences in trajectories); from 23.2% to 5.2% (absolute difference, −18.1%) for accountable justification (difference in differences, −7.0% [95% CI, 9.1% to −2.9%]; *P* < .001); and from 19.9% to 3.7% (absolute difference, −16.3%) for peer comparison (difference in differences, −5.2% [95% CI, −6.9% to −1.6%]; *P* < .001) (**Figure 2**; full regression results available in eTable 3 in Supplement 2).

There were no statistically significant interactions between interventions, and the fully interacted model failed to improve goodness-of-fit over the main effects model (eTable 4A and 4B in Supplement 2). The results of sensitivity analyses treating the entire 18-month baseline and intervention periods as dummy variables (eTable 5 in Supplement 2) were similar to the main analysis (Figure 2; eTable 3 in Supplement 2). Unadjusted qualifying visit counts and prescribing rates before and during the intervention for each study group are available in **Table 2**.

Diagnosis Shifting
Relative to control, none of the interventions was statistically significantly associated with changes over time in the proportion of antibiotic-appropriate acute respiratory tract infection diagnoses among visits for any acute respiratory tract infection diagnosis (ie, there was no evidence of diagnosis shifting; full regression results available in eTable 6 in Supplement 2).

Safety
The rate of return visits for possible bacterial infections within 30 days following visits for acute respiratory tract infection (both antibiotic-inappropriate and potentially antibiotic-appropriate) in which antibiotics were not prescribed was 0.43% (95% CI, 0.25% to 0.70%) among control practices. There was a statistically significantly higher rate of such return visits in the accountable justification plus peer comparison group (1.41% [95% CI, 1.06% to 1.85%]) (eTable 7 in Supplement 2). No other intervention group (including the group applying all 3 interventions simultaneously) had a statistically significantly higher rate of such return visits.

Among return visits, study physicians and an independent data safety and monitoring board reviewed a random sample of 33 cases across all study groups. For 12 cases, it

Figure 1. Flow of Participants in a Trial of Antibiotic Prescribing
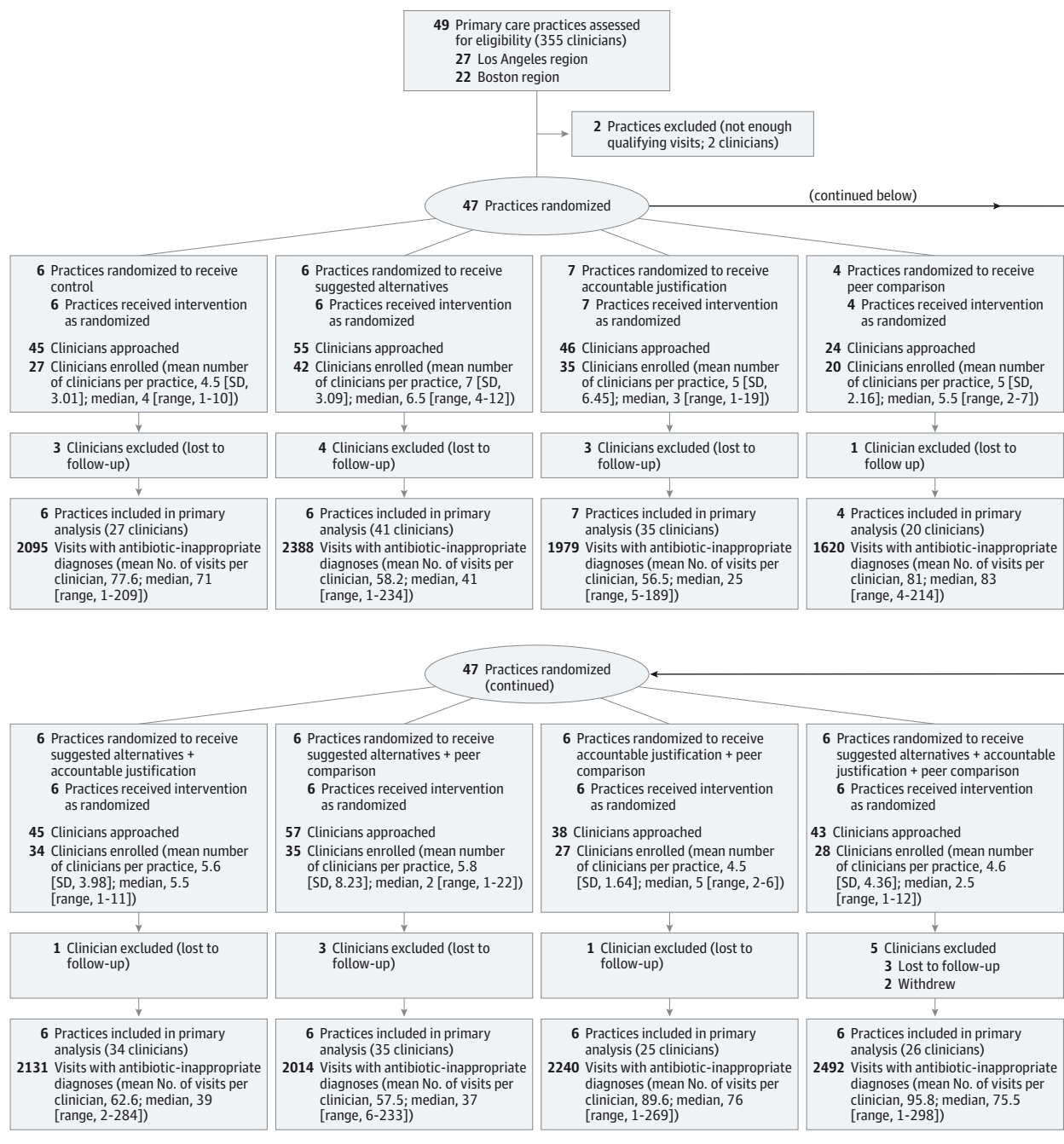
was unlikely that an antibiotic would have been helpful if prescribed at the index visit (eg, "cold symptoms" with clear chest and no fever at return visit); for 8 cases, there was uncertainty (eg, patient returned with a diagnosis of pneumonia, but no chest radiograph was obtained at the index or return visit); and for 13 cases, antibiotics might have been helpful (eg, a patient with influenza diagnosis at initial visit was hospitalized for pneumonia 1 week later; return-visit diagnoses associated with these 13 cases are available in eTable 8 in Supplement 2).

Discussion

We designed, implemented, and evaluated 3 behavioral interventions to curtail overuse of antibiotics for acute respiratory tract infections. We found that 2 socially motivated interventions—accountable justification and peer comparison—resulted in statistically significant reductions in inappropriate antibiotic prescribing, while suggested alternatives, which lacked a social component, had no statistically significant

Table 1. Sample Characteristics During the 18-Month Intervention Period

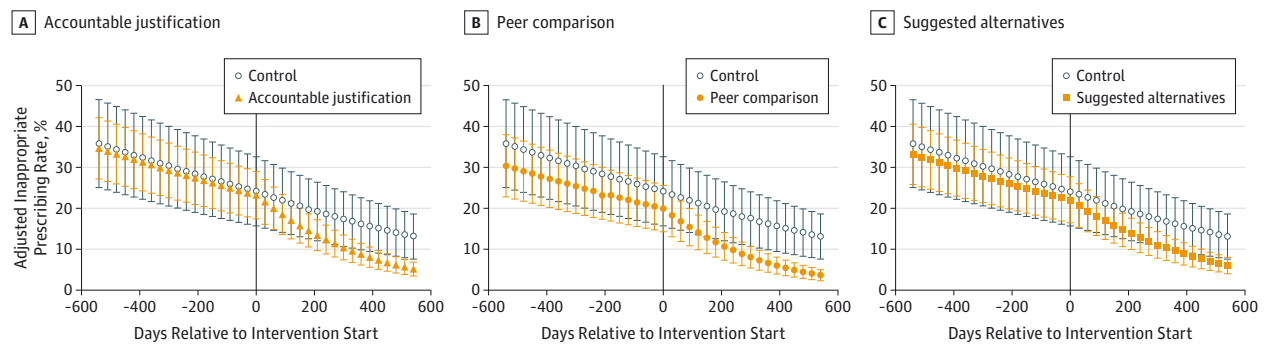|	Overall	Control	Suggested Alternatives	Accountable Justification	Peer Comparison	Suggested Alternatives + Accountable Justification	Suggested Alternatives + Peer Comparison	Accountable Justification + Peer Comparison	Suggested Alternatives + Accountable Justification + Peer Comparison
No. of practices	47	6	6	7	4	6	6	6	6
Clinician Characteristics									
No. of clinicians, No. (%)	248	27 (11)	42 (17)	35 (14)	20 (8)	34 (14)	35 (14)	27 (11)	28 (11)
Age, mean (SD), y	48 (10)	47 (12)	49 (10)	49 (9)	46 (9)	48 (9)	51 (11)	48 (10)	46 (9)
Men, No. (%)	91 (37)	14 (52)	12 (29)	13 (37)	8 (40)	10 (29)	10 (29)	12 (44)	12 (46)
Time since first licensure, mean (SD), y	20 (10)	21 (11)	21 (10)	21 (10)	18 (7)	20 (9)	20 (11)	19 (11)	19 (10)
Time since first licensure, median (IQR), y	18 (12.5-27)	16 (13-30)	20.5 (14-29)	21 (14-28)	17.5 (12-24)	17 (12-26)	18 (12-29)	16 (12-31)	18.5 (13-27)
Clinician type, No. (%)									
Internal Medicine	150 (60)	17 (62)	24 (57)	27 (77)	10 (50)	21 (62)	25 (71)	13 (48)	13 (50)
Family Medicine	32 (13)	3 (11)	5 (12)	1 (3)	6 (30)	5 (15)	2 (6)	6 (22)	4 (15)
Other PCP	16 (6)	2 (7)	5 (12)	1 (3)	0	1 (3)	0	4 (15)	3 (12)
Physician assistant or nurse practitioner	48 (19)	5 (19)	8 (19)	6 (17)	4 (20)	7 (20)	8 (23)	4 (15)	6 (23)
Per-clinician FTE during intervention									
Mean (SD)[a]	0.66 (0.32)	0.80 (0.24)	0.64 (0.38)	0.62 (0.30)	0.80 (0.19)	0.66 (0.29)	0.50 (0.27)	0.79 (0.24)	0.54 (0.38)
Median (IQR)	0.67 (0.41-0.97)	0.84 (0.68-1.0)	0.69 (0.29-1.0)	0.58 (0.42-0.84)	0.80 (0.67-1.0)	0.68 (0.49-0.87)	0.50 (0.32-0.59)	0.85 (0.56-1.0)	0.50 (0.18-0.90)
No. of clinicians by region, No. (%)									
Massachusetts	171 (69)	18 (67)	27 (64)	29 (83)	12 (60)	27 (79)	31 (89)	13 (48)	14 (50)
Southern California	77 (31)	9 (33)	15 (36)	6 (17)	8 (40)	7 (21)	4 (11)	14 (52)	14 (50)
Patient Characteristics									
Age, mean (SD), y	48 (17)	49 (15)	48 (17)	53 (17)	47 (16)	50 (17)	48 (17)	49 (17)	43 (17)
Men, No. (%)	5567 (33)	727 (35)	675 (28)	830 (42)	511 (32)	617 (29)	631 (31)	753 (34)	823 (33)
Race, No. (%)[b]									
White	14 415 (87)	1770 (88)	1990 (86)	1759 (90)	1347 (85)	1796 (85)	1740 (86)	1995 (91)	2018 (84)
Black	834 (5)	78 (4)	118 (5)	96 (5)	100 (6)	117 (6)	158 (8)	49 (2)	118 (5)
Asian	767 (5)	74 (4)	83 (4)	50 (3)	97 (6)	145 (7)	56 (3)	87 (4)	175 (7)
Other race/unknown	516 (3)	94 (4)	109 (5)	48 (2)	39 (3)	47 (2)	56 (3)	49 (3)	83 (4)
Hispanic or Latino ethnicity	5383 (32)	738 (35)	721 (30)	502 (25)	418 (26)	408 (19)	817 (41)	850 (38)	929 (37)
Insurance type, No. (%)[b]									
Medicare	2340 (14)	255 (13)	336 (15)	377 (19)	187 (12)	358 (17)	250 (13)	334 (16)	243 (10)
State or county subsidized[c]	3894 (24)	543 (27)	518 (24)	211 (11)	283 (18)	455 (22)	542 (27)	691 (32)	651 (27)
Private	9737 (59)	1207 (60)	1285 (59)	1276 (65)	1036 (68)	1245 (59)	1147 (57)	1092 (51)	1449 (59)
Self-pay or other	414 (3)	22 (1)	34 (2)	99 (5)	27 (2)	50 (2)	58 (3)	30 (1)	94 (4)
No. of qualifying visits	16 959	2095	2388	1979	1620	2131	2014	2240	2492

Abbreviations: FTE, full-time equivalent; IQR, interquartile range; PCP, primary care practice.

[a] Intervention full-time equivalent was calculated using number of hours in clinic per week during the 18-month intervention period.

[b] Cell counts may not sum to total because of missing values. Covariates were missing in approximately 3% of observations.

[c] Includes Medicaid and other public payers.

Figure 2. Adjusted Rates of Antibiotic Prescribing at Primary Care Office Visits for Antibiotic-Inappropriate Acute Respiratory Tract Infections Over Time

Prescribing rates for each intervention are marginal predictions from hierarchical regression models of intervention effects, adjusted for concurrent exposure to other interventions and clinician and practice random effects. Error bars indicate 95% CIs. Model coefficients are available in eTable 3 in Supplement 2.

Table 2. Unadjusted Visit Counts and Antibiotic Prescribing Rates for Antibiotic-Inappropriate Acute Respiratory Tract Infections During the Baseline and Intervention Periods, by Study Group

	Control	Suggested Alternatives	Accountable Justification	Peer Comparison	Suggested Alternatives + Accountable Justification	Suggested Alternatives + Peer Comparison	Accountable Justification + Peer Comparison	Suggested Alternatives + Accountable Justification + Peer Comparison
Baseline Period								
Visits for antibiotic-inappropriate acute respiratory tract infections, No.	1866	2132	1491	1236	1977	1511	2362	2178
Inappropriate antibiotic prescriptions, No.	692	1057	497	433	702	368	782	558
Rate of inappropriate antibiotic prescribing, % (95% CI)[a]	37.1 (34.9-39.3)	49.6 (47.5-51.7)	33.3 (30.9-37.7)	35.0 (32.4-37.7)	35.5 (33.4-37.6)	24.4 (22.2-26.5)	33.1 (31.2-35.0)	25.6 (23.8-37.5)
Intervention Period								
Visits for antibiotic-inappropriate acute respiratory tract infections, No.	2095	2388	1979	1620	2131	2014	2240	2492
Inappropriate antibiotic prescriptions, No	502	722	324	311	341	139	340	249
Rate of inappropriate antibiotic prescribing, % (95% CI)	24.0 (22.1-25.8)	30.2 (28.4-32.1)	16.4 (14.7-18.0)	19.2 (17.3-21.1)	16.0 (14.5-17.6)	6.9 (5.8-8.0)	15.2 (13.7-16.7)	10.0 (8.8-11.2)

[a] Confidence intervals are exact 95% CIs based on the cumulative density function of the binomial distribution.

effect. There were no statistically significant interactions between interventions; therefore, applying these interventions simultaneously might have additive effects on antibiotic prescribing.

The intervention effects that we observed represent reductions in inappropriate prescribing beyond those attributable to an educational module or observation alone (the Hawthorne effect), which were both applied in the control condition. We believe these effect sizes (5.2 to 7.0 percentage points) are clinically significant, especially when measured against control clinicians who were motivated to join a trial,

knew they were being monitored, and who had relatively low antibiotic prescribing rates at baseline.[32-34]

Insights from behavioral science may have enhanced the effectiveness of interventions. Our peer comparison intervention performed favorably in comparison with traditional audit-and-feedback.[35] Previous studies have shown modest effects[36] or null effects[37] when justification interventions lacked public accountability (ie, lacked mechanisms to publicly display clinicians' written justifications). In contrast, our justification intervention included such a mechanism: the "antibiotic justification note."

All 3 interventions involved modest changes to the practice environment (ie, "nudges"); none restricted clinicians' choice of treatment or changed how clinicians were paid. For some primary care practices, the peer comparison intervention might be the simplest and most pragmatic of these interventions, since it requires no modification of the EHR. However, the peer comparison intervention depends on producing valid performance measures, which can be challenging when data are unavailable or individual-clinician sample sizes are small. The accountable justification intervention, although it requires EHR modification, does not have these drawbacks and can, in theory, be applied to any clinical decision (since clinicians should always be able to articulate a decision-making rationale).

There was little evidence of potential harm associated with any of the 3 interventions. Although the study group applying both the accountable justification and peer comparison interventions had a modestly higher rate of return visits for diagnoses of concern (approximately 1.4% of visits, compared with 0.4% in the control group), no other intervention group (including the group applying all 3 interventions simultaneously) was associated with higher rates of such return visits. We did not assess potential harms associated with prescribing antibiotics unnecessarily; such harms are most likely when inappropriate prescribing rates are highest (ie, when no intervention is applied). However, as with other efforts to change clinical practice, continual monitoring of patient outcomes is advisable.

Our study has limitations. First, the number of clinicians within each cluster was small. Although a high proportion of invited clinicians chose to participate, some did not, which may limit generalizability. Similarly, trial findings might not generalize to primary care practices dissimilar to those enrolled. Second, results are dependent on EHR and billing data, which are imperfect for performance measurement—although in the present context they have demonstrated validity.[38]

Third, logistic and other nonlinear regression methods may lead to effect sizes biased toward the null in randomized studies when important covariates are omitted.[39,40] Fourth, although interactions were not significant, combining nudges in some cases could have attenuated the effects we observed (eg, if prescribing rates hit a "floor"). Therefore, our results might underestimate the independent effects of each individual intervention.

Fifth, our safety analyses were limited to return visits to the clinical organizations studied (possibly underestimating intervention risks) and did not investigate harms caused by unnecessary antibiotics (possibly underestimating risks of *not* applying the interventions). Sixth, we did not distinguish between narrower- and broader-spectrum antibiotics.

Seventh, we did not measure heterogeneity of intervention effects by type of practice, clinician, or EHR product. Eighth, we could not directly measure differences in clinicians' coding habits between types of practice settings. Ninth, persistence of effects is unknown. Last, elements of our analytic approach were determined post hoc, based on guidance during the peer review process. Effect estimates from these post hoc analyses had the same direction and statistical significance as our prespecified analyses.

Conclusions

Among primary care practices, the use of accountable justification and peer comparison as behavioral interventions resulted in lower rates of inappropriate antibiotic prescribing for acute respiratory tract infections.

REFERENCES

1. Tackling a Crisis for the Future Health and Wealth of Nations. Review on Antimicrobial Resistance website. http://amr-review.org/. December 2014. Accessed January 14, 2016.

2. Office of the Press Secretary, The White House. Executive Order—Combating Antibiotic-Resistant Bacteria. Whitehouse.gov website. https://www.whitehouse.gov/the-press-office/2014/09/18/executive-order-combating-antibiotic-resistant-bacteria. 2014. Accessed January 14, 2016.

3. Steinman MA, Gonzales R, Linder JA, Landefeld CS. Changing use of antibiotics in community-based outpatient practice, 1991-1999. *Ann Intern Med*. 2003;138(7):525-533.

4. Grijalva CG, Nuorti JP, Griffin MR. Antibiotic prescription rates for acute respiratory tract infections in US ambulatory settings. *JAMA*. 2009; 302(7):758-766.

5. Barnett ML, Linder JA. Antibiotic prescribing for adults with acute bronchitis in the United States, 1996-2010. *JAMA*. 2014;311(19):2020-2022.

6. Fairlie T, Shapiro DJ, Hersh AL, Hicks LA. National trends in visit rates and antibiotic prescribing for adults with acute sinusitis. *Arch Intern Med*. 2012;172(19):1513-1514.

7. Linder JA. Antibiotic Prescribing for acute respiratory infections—success that's way off the mark: comment on "A cluster randomized trial of decision support strategies for reducing antibiotic use in acute bronchitis". *JAMA Intern Med*. 2013;173 (4):273-275.

8. Ranji S, Steinman MA, Shojania KG, et al. Antibiotic Prescribing Behavior: Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies. Rockville, MD: Agency for Healthcare Research and Quality; January 2006. Technical Review 9.

9. Houle SK, McAlister FA, Jackevicius CA, Chuck AW, Tsuyuki RT. Does performance-based remuneration for individual health care practitioners affect patient care?: a systematic review. *Ann Intern Med*. 2012;157(12):889-899.

10. Jha AK, Joynt KE, Orav EJ, Epstein AM. The long-term effect of premier pay for performance on patient outcomes. *N Engl J Med*. 2012;366(17): 1606-1615.

11. Embi PJ, Leonard AC. Evaluating alert fatigue over time to EHR-based clinical trial alerts: findings from a randomized controlled study. *J Am Med Inform Assoc*. 2012;19(e1):e145-e148.

12. Gerber JS, Prasad PA, Fiks AG, et al. Effect of an outpatient antimicrobial stewardship intervention on broad-spectrum antibiotic prescribing by primary care pediatricians: a randomized trial. *JAMA*. 2013;309(22):2345-2352.

13. Metlay JP, Camargo CA Jr, MacKenzie T, et al; IMPAACT Investigators. Cluster-randomized trial to improve antibiotic use for adults with acute respiratory infections treated in emergency departments. *Ann Emerg Med*. 2007;50(3):221-230.

14. Office of the Press Secretary, The White House. Executive Order—Using Behavioral Science Insights to Better Serve the American People. Whitehouse.gov website. https://www.whitehouse.gov/the-press-office/2015/09/15/executive-order-using-behavioral-science-insights-better-serve-american. September 15, 2015. Accessed January 14, 2016.

15. Loewenstein G, Brennan T, Volpp KG. Asymmetric paternalism to improve health behaviors. *JAMA*. 2007;298(20):2415-2417.

16. Observational Medical Outcomes Partnership. OMOP Specifications for Implementation of Standard Vocabularies in Observational Data Analysis Version 3.0. http://www.humangenome.org/zeynep/OMOP%20Specifications%20for%20Implementation%20of%20%20Vocabularies%20V3.0_June%2023-2010.pdf. June 2010. Accessed January 14, 2016.

17. Cooper RJ, Hoffman JR, Bartlett JG, et al; American Academy of Family Physicians; American College of Physicians-American Society of Internal Medicine; Centers for Disease Control. Principles of appropriate antibiotic use for acute pharyngitis in adults: background. *Ann Intern Med*. 2001;134(6): 509-517.

18. Hickner JM, Bartlett JG, Besser RE, Gonzales R, Hoffman JR, Sande MA; American Academy of Family Physicians; American College of Physicians-American Society of Internal Medicine; Centers for Disease Control; Infectious Diseases Society of America. Principles of appropriate antibiotic use for acute rhinosinusitis in adults: background. *Ann Intern Med*. 2001;134(6):498-505.

19. Grice H. *Some further notes on logic and conversation*. Cambridge, MA: Harvard University Press; 1978.

20. Lerner JS, Tetlock PE. Accounting for the effects of accountability. *Psychol Bull*. 1999;125(2): 255-275.

21. Baumeister RE, Hutton DG. Self-presentation theory: self-construction and audience pleading. In: Mullen B, Goethals G, eds. *Theories of Group Behavior*. New York, NY: Springer Verlag; 1987:71-87

22. Jerdee T, Rosen B. Effects of opportunity to communicate and visibility of individual decisions on behavior in the common interest. *J Appl Psychol*. 1974;59(6):712-716.

23. Milinski M, Semmann D, Krambeck HJ. Reputation helps solve the "tragedy of the commons." *Nature*. 2002;415(6870):424-426.

24. Cialdini R, Reno R, Kallgren C. A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *J Pers Soc Psychol*. 1990;58(6):1015-1026.

25. De Cremer D, Barker M. Accountability and cooperation in social dilemmas: the influence of others' reputational concerns. *Curr Psychol*. 2003; 22(2):155-163.

26. Goldstein N, Cialdini R, Griskevicius V. A room with a viewpoint: using social norms to motivate environmental conservation in hotels. *J Consum Res*. 2008;35(3):472-482.

27. Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. *JAMA*. 2001;285(22):2871-2879.

28. Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V. The constructive, destructive, and reconstructive power of social norms. *Psychol Sci*. 2007;18(5):429-434.

29. Gunst RF, Mason RL. Fractional factorial design. *WIREs Comp Stat*. 2009;1:234-244.

30. Wald A. Tests of statistical hypothesis concerning several parameters when the number of observations is large. *Trans Am Math Soc*. 1943;54 (3):426-482.

31. Snow V, Mottur-Pilson C, Gonzales R; American College of Physicians-American Society of Internal Medicine; American Academy of Family Physicians; Centers for Disease Control; Infectious Diseases Society of America. Principles of appropriate antibiotic use for treatment of nonspecific upper respiratory tract infections in adults. *Ann Intern Med*. 2001;134(6):487-489.

32. Gonzales R, Anderer T, McCulloch CE, et al. A cluster randomized trial of decision support strategies for reducing antibiotic use in acute bronchitis. *JAMA Intern Med*. 2013;173(4):267-273.

33. Jones BE, Sauer B, Jones MM, et al. Variation in outpatient antibiotic prescribing for acute respiratory infections in the veteran population: a cross-sectional study. *Ann Intern Med*. 2015;163 (2):73-80.

34. Little P, Stuart B, Francis N, et al; GRACE consortium. Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: a multinational, cluster, randomised, factorial, controlled trial. *Lancet*. 2013;382(9899):1175-1182.

35. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2006;2(2): CD000259.

36. O'Connor SD, Sodickson AD, Ip IK, et al. Journal club: Requiring clinical justification to override repeat imaging decision support: impact on CT use. *AJR Am J Roentgenol*. 2014;203(5):W482-W490.

37. Procop GW, Keating C, Stagno P, et al. Reducing duplicate testing: a comparison of two clinical decision support tools. *Am J Clin Pathol*. 2015;143 (5):623-626.

38. Linder JA, Bates DW, Williams DH, Connolly MA, Middleton B. Acute infections in primary care: accuracy of electronic diagnoses and electronic antibiotic prescribing. *J Am Med Inform Assoc*. 2006;13(1):61-66.

39. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*. 1984;71(3):432-444.

40. Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. *Int Stat Rev*. 1991;58(2):227-240.