Failure of Automated Telephone Outreach With Speech Recognition to Improve Colorectal Cancer Screening

A Randomized Controlled Trial

Steven R. Simon, MD, MPH; Fang Zhang, PhD; Stephen B. Soumerai, ScD; Arthur Ensroth, MPH; Lydia Bernstein, MPH; Robert H. Fletcher, MD, MSc; Dennis Ross-Degnan, ScD

Background: Automated telephone outreach with speech recognition (ATO-SR) is used extensively by health plans. Whether ATO-SR can increase rates of colorectal cancer (CRC) screening is unknown.

Methods: We randomly allocated 40,000 health plan members to ATO-SR and 40,000 to usual care, of whom 10,432 and 10,506 in the intervention and usual care groups, respectively, had not been previously screened and were therefore eligible for analysis. The intervention was a single interactive outreach call using speech recognition to engage participants in conversation about the importance of CRC screening and options for and barriers to screening. The intervention directed participants to contact their primary care provider to schedule screening. The primary end point was any CRC screening in the year following intervention. Colonoscopy in the year following intervention was a secondary outcome.

Results: The incidence of any CRC screening was 30.6% in the intervention group and 30.4% in the usual care group (P = .76). After adjustment for available covariates, there remained no intervention effect (adjusted odds ratio [OR], 1.01; 95% confidence interval [CI], 0.94-1.07). A total of 21.4% of members in the intervention group and 20.3% in the usual care group underwent colonoscopy (P = .04). In multivariate analysis, there was a small intervention effect on colonoscopy (OR, 1.08; 95% CI, 1.00-1.16).

Conclusions: This study showed that ATO-SR failed to improve rates of CRC screening. Future studies should examine approaches that combine efforts to target patients and their health care providers to overcome the barriers to CRC screening.

Trial Registration: clinicaltrials.gov Identifier: NCT00792285

Arch Intern Med. 2010;170(3):264-270
consistent with several conceptual frameworks for improving health care quality, including Wagner's chronic care model, and although as a single intervention it may represent a weak stimulus to behavior change. Some of the largest health plans and delivery systems in the United States, such as Kaiser Permanente and Blue Cross Blue Shield, have widely adopted this technology for a variety of outreach efforts, ranging from increasing immunization rates for influenza to encouraging at-risk patients to undergo osteoporosis screening. One uncontrolled study suggested that this technology increases use of preventive health services, including mammograms and Papanicolaou tests, but to our knowledge, to date no randomized controlled trials (RCTs) have assessed ATO to improve cancer screening. Live telephone outreach and reminders have been shown to increase rates of screening for other conditions, such as cervical cancer and breast cancer, although human involvement in the outreach process makes these approaches impractical for large populations.

Our health plan embarked on a program of ATO-SR to increase CRC screening. Because limited resources prevented the organization from targeting all eligible individuals with the initial intervention, the health plan randomly allocated the intervention to a predetermined number of age-eligible members. This quality improvement program therefore resulted in a large, naturally occurring RCT with more than 20,000 eligible patients to evaluate the effectiveness of ATO-SR to increase rates of CRC screening. Evaluation of this pragmatic clinical trial provided an opportunity to assess the effectiveness of a system-wide intervention as it was implemented under usual conditions, to provide information for health plans considering similar interventions.

STUDY DESIGN

The study was an RCT conducted at Harvard Pilgrim Health Care, a health plan serving approximately 1 million members in Massachusetts, New Hampshire, and Maine. The study used a randomized, nonblinded, 2-group design comparing an ATO intervention with usual care. Because the intervention was initially conducted as an unfunded quality improvement initiative of the health plan, there was no requirement to obtain informed consent, thus eliminating problems of generalizability owing to volunteer bias. The Harvard Pilgrim Health Care human studies committee determined that the study was exempt from formal institutional review board approval under federal regulations. The trial was registered at ClinicalTrials.gov.

A flowchart of study participants is shown in the Figure. We identified all commercially insured (ie, non-Medicare and non-Medicaid) men and women aged 50 to 64 years prior to the intervention (N = 160,285) and excluded 50,497 individuals for the following reasons: participation in a similar outreach program in the previous year (n = 34,817), prior notification of preference not to be included in quality improvement programs (n = 1815), enrollment in existing Harvard Pilgrim telephone-based disease management programs (n = 5797), and absence of a telephone number in the membership records (n = 11,977). We examined the residential addresses of the remaining 109,788 men and women and found 26,666 pairs of members sharing the same household. We randomly selected and excluded 1 member from each pair to prevent contamination, leaving a sample of 83,122 members. Health plan resources were sufficient to include only 40,000 members in the ATO. We therefore randomly selected 80,000 members whom we randomly allocated to intervention and usual care arms, using a computerized random-number generator. The health plan included all members meeting these criteria, without determining whether the members were currently due for CRC screening, because the ATO delivered an educational message that was designed to be relevant whether or not members had been previously screened.

INTERVENTION

Because this was an operational program, the intervention format of a single ATO call was determined by the health plan. As much as possible, we designed the ATO call to reflect the insights of several theories of behavior change, including the General Model of the Determinants of behavioral change, a synthesis of behavioral theories. This model highlights the importance of patients' knowledge, attitudes, and self-efficacy in achieving intended behavior change outcomes. The intervention primarily targeted knowledge deficits but also attempted to address attitudes and self-efficacy by providing a positive and encouraging message as well as a clear and attainable next step for action. If completed by the target members, the ATO calls assessed awareness of and provided information about the importance of CRC screening, different options for screening, and family history as a risk factor for CRC. To enhance self-efficacy, the calls provided positive reinforcement for participants who either reported that they had been screened in the past or that they were considering screening in the future; for individuals who indicated that they were not planning to undergo screening, the language was nonjudgmental and emphasized the universal need for screening and the options available. For all participants, the intervention provided a consistent message that calling their physician was the next step to arrange screening.

The ATO intervention ran from March to May 2005. Building on our theoretical framework, we developed a script and branching algorithm for the ATO call to members to educate them about the risk of CRC and about the importance and methods of screening, and to encourage them to contact their primary care providers to arrange for CRC screening. Using speech recognition technology, the call attempted to engage members in conversation after careful identification of the individual answering the telephone either as the intended recipient or as someone who could take a message for the intended member. If a member was not reached directly with outbound calling, the system left a message inviting the member to call back, at which time the system implemented the intervention script. The ATO system detected the speech of the participant and, using a computer algorithm, responded with prerecorded human conversation appropriate to the words articulated by the participant. Most participants seemed to recognize that they were not speaking to a "live" person within a sentence or 2 of interaction. After verifying identity and securing the member's permission to proceed with the call interaction, the system engaged the member in a conversation about CRC screening, including the following topics:

- Ascertaining whether the member had previously undergone CRC screening or had plans to do so.
- Describing the prevalence of CRC and the potential for screening to reduce disease burden.
- Offering to provide information on the methods of CRC screening.
- Offering to mail additional information on CRC screening to the member (based on materials produced by the Cen-
OUTCOME MEASURES

The main outcome measure was screening for CRC within 12 months following the intervention. The CRC screening included any of the recommended screening modalities, including fecal occult blood testing, double-contrast barium enema, flexible sigmoidoscopy, or colonoscopy. As a secondary outcome measure, we examined screening by colonoscopy during the 12-month period following the intervention; we specifically included colonoscopy as a secondary outcome because of increased public awareness of colonoscopy as the most sensitive and frequently recommended screening test, and the possibility that our direct-to-patient intervention might have effects on rates of this particular screening modality. We also included time to any CRC screening and time to colonoscopy as additional secondary outcome measures. We used the presence of Current Procedural Terminology or International Classification of Diseases, Ninth Revision (ICD-9) codes (see eTable 1 and eTable 2) in the administrative claims database of the health plan to ascertain the outcomes.

SAMPLING

Because the outreach was implemented as an operational quality improvement intervention, it included 351 individuals who had previously notified Harvard Pilgrim of their preference to be excluded from research. Therefore, we excluded these individuals (160 in the intervention arm and 191 in the usual care arm) from participation in the usual care arm.

Figure. A flowchart of study participants. CRC indicates colorectal cancer. *Participants could have multiple reasons for exclusion. †Health plan members had previously notified Harvard Pilgrim of their preference to be excluded from research. Therefore, we excluded these individuals from participation in the usual care arm. **Receipt of intervention was defined as the intended participant confirming his or her identity and giving consent to the automated system to proceed with the discussion. ††There were incomplete data to make outreach call. The call answered but recipient hung up before identification. §This includes 119 nonparticipants who declined to leave a message for study participants. ¶**Health plan members had previously notified Harvard Pilgrim of their preference to be excluded from research. Therefore, we excluded these individuals from participation in the usual care arm. ¶¶Continuous health plan enrollment for the 12 months preceding and the 12 months following the intervention was required. §§Based on review of administrative health plan data for the 5-year period preceding intervention. §§§Based on presence of administrative data indicating completion of colonoscopy, sigmoidoscopy, or double-contrast barium enema in the 5 years preceding intervention or fecal occult blood testing in the year preceding intervention.

©2010 American Medical Association. All rights reserved.
arm) from all analyses. It is also important to emphasize that the initial sample did not exclude individuals who were ineligible for CRC screening owing to prior screening according to the recommended protocol, because the intervention itself was able to accommodate them through variations in the interaction. We excluded these individuals from the statistical analysis using the following approach. We first limited the analytic sample to health plan members who had been continuous members during the year preceding the intervention; this criterion increased the likelihood that any previously completed screening for CRC during that period would have been included in the data files. We also limited the sample to individuals with continuous membership during the 12 months following intervention to ensure complete ascertainment of outcomes. These continuous enrollment requirements led to the exclusion of 12,631 individuals in the intervention group and 12,641 individuals in the usual care group. We then excluded any individuals who had evidence of colonoscopy, sigmoidoscopy, or barium enema in the 5 years prior to intervention, or fecal occult blood testing in the year prior to intervention, because these individuals would not have been due for screening at the time of the intervention. We also excluded patients who had evidence of adenomatous polyps or CRC diagnosed in the 5 years preceding the intervention, because these individuals (5,566 in the intervention group and 5,471 in the usual care group) would have been subject to different recommendations for screening or surveillance than the general population. These exclusions reduced the analytic sample to 10,432 members in the intervention group and 10,506 in the usual care group.

### Statistical Analysis

The goal of the primary analysis was to test the hypothesis that an ATO intervention would increase rates of CRC screening. The analyses followed intention-to-treat principles. We identified covariates potentially associated with receipt of CRC screening before analysis, including age, sex, and diagnosis of inflammatory bowel disease during the baseline period, hospitalization during the baseline period, and socioeconomic status. We used health plan membership records to determine the age and sex of participants. We used the home address of the individual, mapped to the census block, and 2000 US Census data to determine the corresponding average neighborhood household income and racial composition and applied these to the individual. We evaluated differences between the ATO intervention and the usual care groups in the year following the intervention for the primary end point of any CRC screening and the secondary end point of colonoscopy using logistic regression. For the secondary analyses of time-to-event (any CRC screening and colonoscopy), we used Kaplan-Meier estimates compared by the log-rank test. We also used proportional hazard models that controlled for our other covariates in assessing time to screening; proportional hazards assumptions were tested and confirmed. We used SAS statistical software (version 9.1; SAS Institute Inc, Cary, North Carolina) for all analyses.

### RESULTS

#### PATIENT CHARACTERISTICS

A total of 20,938 members were eligible for analysis: 10,432 in the intervention group and 10,506 in the usual care group (Table). There were no baseline differences between the 2 study groups on any of the measured variables. Among the 40,000 health plan members allocated to the intervention group, 24,488 (61%) received the intervention as evidenced by their verbal interaction with the ATO system.

### PRIMARY OUTCOME MEASURE

The incidence of the primary end point, any CRC screening in the year following the intervention, was 30.6% in the intervention group and 30.4% in the usual care group (P = .76). After adjustment for available covariates, we did not observe any intervention effect (adjusted odds ratio [OR], 1.01; 95% confidence interval [CI], 0.94-1.07).

### SECONDARY OUTCOME MEASURES

In the analysis of the secondary end point, a total of 21.4% of members in the intervention group and 20.3% in the usual care group underwent colonoscopy following the intervention (P = .04). In multivariate analysis, there was a small effect of the intervention on the adjusted odds of receiving a colonoscopy (OR, 1.08; 95% CI, 1.00-1.16).

In survival analysis, there was no statistical difference between the intervention and usual care groups in time to any CRC screening (P = .61 by log-rank test). In comparison, the time to colonoscopy in the intervention group was slightly less than the time for the usual care group (P = .03). The corresponding multivariate survival analyses yielded adjusted hazard ratios of 1.01 (95% CI, 0.96-1.06) for time to any CRC screening and 1.07 (95% CI, 1.01-1.14) for time to colonoscopy.

### COMMENT

Although ATO-SR programs have been widely disseminated throughout the United States to promote a variety of health behaviors, in this large RCT among more than 20,000 eligible health plan members who had not al-

### Table. Baseline Characteristics of Study Participants Eligible for Analysis

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Intervention Group (n=10,432)</th>
<th>Usual Care Group (n=10,506)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>56.7 (4.1)</td>
<td>56.7 (4.1)</td>
<td>.95</td>
</tr>
<tr>
<td>Women, %</td>
<td>53.3</td>
<td>52.4</td>
<td>.20</td>
</tr>
<tr>
<td>Inflammatory bowel disease, %</td>
<td>0.47</td>
<td>0.62</td>
<td>.14</td>
</tr>
<tr>
<td>Prior hospitalization, %</td>
<td>21.5</td>
<td>21.8</td>
<td>.63</td>
</tr>
<tr>
<td>Race (US census derived)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>4.7 (12.8)</td>
<td>4.5 (12.4)</td>
<td>.43</td>
</tr>
<tr>
<td>White</td>
<td>85.8 (19.9)</td>
<td>85.9 (19.4)</td>
<td>.67</td>
</tr>
<tr>
<td>Other</td>
<td>9.6 (11.5)</td>
<td>9.6 (11.5)</td>
<td>.99</td>
</tr>
<tr>
<td>Income below FPL (US census derived)</td>
<td></td>
<td></td>
<td>.76</td>
</tr>
</tbody>
</table>

Abbreviation: FPL, federal poverty level.

a Eligibility criteria, as outlined in the Figure and the “Methods” section, included not having undergone screening in the 5-year period preceding the intervention. Data are presented as means (SDs) except where noted.

b We used the home address of the individual, mapped to the census block, and 2000 US Census data to determine the corresponding average neighborhood household income and racial composition and applied these to the individual.

©2010 American Medical Association. All rights reserved.
ready been screened, we found that this intervention did not improve rates of CRC screening. Although there was a small increase in the rate of colonoscopy in the intervention group and some acceleration in the time to colonoscopy among participants receiving the intervention, overall this approach did not result in clinically meaningful improvements in CRC screening.

Prior studies provided evidence to support ATO interventions to improve quality of care,\textsuperscript{28,40} including previous studies suggesting the possibility that analogous strategies may improve rates of screening for other cancers.\textsuperscript{34,41} Unfortunately, both Crawford et al\textsuperscript{32} and Corkrey et al\textsuperscript{34} used nonrandomized study designs that were susceptible to selection bias as an alternative explanation for the apparent observed effect of the intervention.

Several factors may account for why the intervention in our study did not improve rates of CRC screening. First, behavioral theory suggests that a single computer-generated telephone call unaccompanied by other supporting and reinforcing components constitutes a weak stimulus for behavior change.\textsuperscript{42} Even when the call components are designed to target specific knowledge-related barriers and to enhance self-efficacy, this type of intervention may be insufficient to achieve measurable impacts on CRC screening. Furthermore, unlike some prior studies,\textsuperscript{28,34} this large RCT was performed among an unselected health plan population rather than among volunteers for a clinical trial. The intervention may not have penetrated this unselected population to the extent that it might have among a volunteer population. As such, this trial may give a more accurate representation of the effectiveness of this intervention in a real-world population.

The fact that the intervention did not reach 40% of targeted individuals may also have been a key factor underlying the lack of observed effect. By limiting the analyses to health plan members who had not already been screened, the study focused on a segment of the population that may intrinsically be more resistant to screening interventions. Participants may have received an outreach call shortly after or long before a scheduled physician office visit, delaying their ability to act in a positive way on the recommendations in the call. The outreach calls might have had more impact if they could have been timed to occur shortly before a scheduled visit. Furthermore, this ATO intervention required patients to initiate contact with their physicians to arrange screening; other approaches that do not require this level of patient activation may be more effective.

In addition, there are barriers to CRC screening that are not present, or are present to lesser degrees, for other cancer conditions. For example, many patients are embarrassed or unwilling to talk about CRC with their clinicians, and study participants may have been reluctant to follow the advice of the ATO intervention to contact their physician to discuss screening. Furthermore, even if patients contact their primary care providers’ offices to discuss screening, obtaining screening via colonoscopy, currently considered the gold standard for screening,\textsuperscript{8} involves scheduling via a gastroenterology office, obtaining and completing the necessary preparation, undergoing required preprocedure testing (in some cases, to ensure normal renal function), arranging to have a driver for the procedure because of sedation, and then actually undergoing the procedure that, for some patients, is associated with anxiety or discomfort. Given these additional barriers to screening for CRC, which exceed those associated with mammography or Papanicolaou smears, it may not be surprising that a single ATO intervention, as implemented for hundreds of thousands of health plan members throughout the United States,\textsuperscript{29} did not result in measurable increases in screening rates.

While some conditions may be amenable to improvement with IVR-based technology, it is also worth pointing out that recent evidence suggests that this approach may not work for a range of notable clinical and public health problems. For example, recent studies have failed to show a benefit of IVR interventions to improve asthma care and bone-density screening for osteoporosis.\textsuperscript{43,44} Colorectal cancer screening may be among a set of challenging conditions that are resistant to this type of intervention.

Strengths of this study include its large sample size and inclusion of an unselected, representative sample of health plan members, representing a broadly generalizable population of mostly middle-class individuals with some degree of racial and ethnic diversity. The randomized controlled design enabled the comparison of screening rates in the intervention group with a contemporaneous group who received usual care, thereby accounting for possible secular trends. The adherence to the intention-to-treat principle in the analyses is also a strength of the study; we did not conduct a per-protocol analysis, limited to those in the intervention group who actually received the intervention, because such an analysis would suffer from irreconcilable selection bias.

In addition to the fundamental problem that the intervention reached only 60% of targeted individuals, the study has several other limitations. First, the study was not blinded; as such, it is possible that health plan members in the usual care group became aware of the health plan’s outreach efforts to increase CRC screening through communication with members in the intervention group, thereby dampening the potential effect of the intervention. Moreover, members in the intervention group may have told their clinicians about the ATO calls, resulting in their increased attention to CRC screening for all patients, including those allocated to usual care. Second, the intervention itself was limited to a single ATO call that directed health-plan members to call their physician’s office to discuss and arrange recommended screening for CRC. Although large segments of the population are in need of activation to undergo CRC screening,\textsuperscript{13} this scalable but low-intensity intervention may not have the potency to overcome the many barriers to obtaining CRC screening. Third, the study participants, predominantly white and not poor, all had health insurance through the same health plan; while this nonheterogeneity enhanced internal validity, it also limits generalizability of the results.

Widely generalizable, effective, and affordable interventions are critically needed to improve CRC screening and to bridge other quality chasms in large popula-
tions. However, this particular widely disseminated intervention did not appreciably increase rates of CRC screening, suggesting, once again, an occasion when potentially ineffective policies and programs achieve widespread acceptance before evidence is available from controlled trials. Future studies should address the multiple barriers to CRC screening, engaging patients and their health care providers and facilitating the several steps required to obtain screening, in order to improve rates of this potentially life-saving screening intervention.

Accepted for Publication: September 19, 2009.

Correspondence: Steven R. Simon, MD, MPH, Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, 133 Brookline Ave, Sixth Floor, Boston, MA 02215 (steven_simon@hphc.org).

Author Contributions: Study concept and design: Simon, Ensrooth, Bernstein, Fletcher, and Ross-Degnan. Acquisition of data: Ensrooth. Analysis and interpretation of data: Simon, Zhang, Soumerai, Ensrooth, Fletcher, and Ross-Degnan. Drafting of the manuscript: Simon and Bernstein. Critical revision of the manuscript for important intellectual content: Simon, Zhang, Soumerai, Ensrooth, Fletcher, and Ross-Degnan. Statistical analysis: Zhang and Ross-Degnan. Obtained funding: Simon. Administrative, technical, and material support: Zhang, Soumerai, and Ensrooth.

Financial Disclosure: None reported.

Funding/Support: This work was supported by a grant from the Harvard Pilgrim Health Care Foundation.

Role of the Sponsors: The Harvard Pilgrim Health Care Foundation had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the Harvard Pilgrim Health Care Foundation.

Online-Only Material: eFigures 1 and 2 and eTables 1 and 2 are available at http://www.archinternmed.com.

Additional Contributions: Ken Kleinman, ScD, provided statistical advice in the design of the study. Christina Kara provided administrative support and assistance with manuscript preparation. The anonymous peer reviewers supplied helpful advice and suggestions for improving the manuscript.

REFERENCES


