



Published in final edited form as:

Med Care. 2017 August ; 55(8): 797–805. doi:10.1097/MLR.0000000000000759.

Randomized Trial of Reducing Ambulatory Malpractice and Safety Risk: Results of the Massachusetts PROMISES Project

Gordon D. Schiff, MD^{1,2,*}, Harry Reyes Nieva, BA^{1,2,3}, Paula Griswold, MS⁴, Nicholas Leydon, MPH, MBA^{5,6}, Judy Ling, BA⁵, Frank Federico, RPh⁷, Carol Keohane, MS, RN⁸, Bonnie Rae Ellis, BSN, RN⁹, Cathy Foskett, BSN, RN¹, E. John Orav, PhD^{1,2,10}, Catherine Yoon, MS¹, Don Goldmann, MD^{2,7,10}, Joel S. Weissman, PhD^{2,10,11}, David W. Bates, MD, MSc^{1,2,10}, Madeleine Biondolillo, MD⁵, and Sara J. Singer, PhD, MBA^{2,10,12}

¹Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital; 1620 Tremont Street, BC-3, Boston, MA 02120; Tel. 617-281-7628; Fax: 617-732-7072

²Harvard Medical School; 1620 Tremont Street, BC-3, Boston, MA 02120; Tel. 617-732-4814; Fax: 617-732-7072

³Johns Hopkins Bloomberg School of Public Health; 1620 Tremont Street, BC-3, Boston, MA 02120; Tel. 617-582-4882; Fax: 617-582-4890

⁴Massachusetts Coalition for the Prevention of Medical Errors; 500 District Avenue, Burlington, MA 01803; Tel. 781-262-6081; Fax: 781-262-6181

⁵Bureau of Health Care Safety and Quality, Massachusetts Department of Public Health; 99 Chauncy St #200, Boston, MA 02111; Tel. 508-361-9640; Fax: N/A

⁶North Shore Medical Center; 81 Highland Avenue, Salem, MA 01970; Tel. 978-354-2094; Fax: 978-740-4368

⁷Institute for Healthcare Improvement; 20 University Road, 7th Floor, Cambridge, MA 02138; Tel. 617-301-4800; Fax: 617-301-4830

⁸Risk Management Foundation of the Harvard Medical Institutions (CRICO); 1325 Boylston Street, Boston, MA 02215; Tel. 617-679-1415; Fax: 617-679-1430

⁹Coverys Corporation; One Financial Center, Boston, MA 02205; Tel. 508-344-5106; Fax: 617-330-6168

¹⁰Harvard T. H. Chan School of Public Health; 1620 Tremont Street, BC-3, Boston, MA 02120; Tel. 617-732-5899; Fax: 617-732-7072

¹¹Center for Surgery and Public Health, Brigham and Women's Hospital; 1620 Tremont Street, BC-4, Boston, MA 02120; Tel. 617-525-7727; Fax: 617-525-7723

¹²Mongan Institute for Health Policy, Massachusetts General Hospital; 677 Huntington Ave, Boston, MA 02115; Tel. 617-432-7139; Fax: 617-432-3699

*Corresponding Author: Gordon D. Schiff, MD; Tel. 617-732-4814; Fax: 617-732-7072; gschiff@partners.org.

Disclosure of potential conflicts of interest: The authors declare they have no conflicts of interest, including relevant financial interests, activities, relationships, or affiliations.

Abstract

Objective—Evaluate application of quality improvement approaches to key ambulatory malpractice risk and safety areas.

Study setting—Twenty-five small-to-medium-sized primary care practices (16 intervention; 9 control) in Massachusetts.

Study design—Controlled trial of a 15-month intervention including exposure to a learning network, webinars, face-to-face meetings and coaching by improvement advisors **targeting “3+1”** high-risk domains: test result, referral, and medication management plus culture/communication issues **evaluated by survey and chart review tools**.

Data collection methods—Chart reviews conducted at baseline and post-intervention for intervention sites. Staff and patient survey data collected at baseline and post-intervention for intervention and control sites.

Principal findings—Chart reviews demonstrated significant improvements in documentation of abnormal results, patient notification, documentation of an action or treatment plan, and evidence of a completed plan (all $p < 0.001$). Mean days between laboratory test date and evidence of completed action/treatment plan decreased by 19.4 days ($p < 0.001$). Staff surveys showed modest but non-significant improvement for intervention practices relative to controls overall and for the three high-risk domains that were the focus of PROMISES.

Conclusions—A consortium of stakeholders, quality improvement tools, coaches and learning network decreased selected ambulatory safety risks often seen in malpractice claims.

Keywords

Primary care; care improvement; satisfaction; patient safety; malpractice

INTRODUCTION

Over the past decade, attention to patient safety and malpractice issues has increasingly focused on ambulatory, particularly primary care, settings.¹⁻⁵ Many ambulatory malpractice claims demonstrate preventable harm and recent studies have suggested that such cases may be less defensible than inpatient claims^{6,7} pointing to significant opportunities for safer care. The ambulatory setting is rife with safety risks related to care characterized by high volumes, increasing production pressures, fragmented often poorly coordinated care, and diagnostic, handoff and health information technology challenges.⁶⁻⁹ Compared to inpatient facilities, ambulatory settings, particularly smaller offices lack safeguards, risk management support, and regulatory oversights.^{5,10} Despite its importance, few rigorously evaluated interventions to improve ambulatory safety have been reported, with most more narrowly on specific domains such as medication errors.¹¹

Guided by the perspective that the best way to reduce malpractice is to address problems that often underlie suboptimal care, the AHRQ-funded PROMISES (Proactive Reduction of Outpatient Malpractice: Improving Safety, Efficiency, and Satisfaction) Project applied quality improvement techniques to key ambulatory patient safety areas.^{12,13,14} The project

was led by a coalition that included the Massachusetts Department of Public Health (DPH), Brigham and Women's Center for Patient Safety Research and Practice, the States' two leading malpractice insurers, the Massachusetts Coalition for the Prevention of Medical Errors, and the Institute for Healthcare Improvement (IHI). PROMISES employed a controlled trial design involving 25 small-to-medium-sized adult primary care practices (16 intervention; 9 control) in Massachusetts. It focused on "3+1" high-risk clinical domains: management of laboratory test results, referrals, and medications, plus overarching culture and communication issues.

An evaluation team based at the Harvard T.H. Chan School of Public Health and Brigham and Women's Hospital worked to evaluate and feed back the effects of the intervention using multimodal quantitative and qualitative approaches including chart reviews, patient and staff surveys, and participant interviews and observations. Baseline results from the staff surveys have been reported previously.¹³ Here we report on the study's main findings from pre- and post-intervention chart reviews, and staff and patient surveys.

METHODS

Study Design and Setting—The PROMISES study evaluated a quality improvement intervention focusing on 25 small-to-medium-sized adult primary care practices in Massachusetts. We used stratified randomization to assign 16 to the intervention arm and 9 to a control group. We tested a collaborative improvement intervention that emphasized training and in-office coaching by improvement advisors as well as shared learning methods to develop, test and implement changes within the "3+1" high-risk domains. Both also received feedback results from baseline data collection at the start of the intervention period.

Eligible study sites were primary care practices with 1–10 physicians and included affiliates of larger healthcare networks and independent providers. We identified practices through collaborating malpractice insurers and invited them to enroll. Of 202 practices assessed for eligibility, 26 did not meet inclusion criteria. Of 176 practices invited to participate, 82 responded to an initial inquiry, and 25 agreed to participate. The remainder declined largely due to lack of time and/or competing priorities. We randomized participating practices into 16 intervention and 9 control sites, stratified to balance practice size and health systems, using a random number generator.

All sites received \$1000 reimbursement to compensate for costs associated with facilitating the staff and patient surveys. Intervention sites received an additional \$3000 to help defray expenses incurred through participation, particularly facilitation of chart reviews. The study was approved by the Massachusetts Department of Public Health and Partners HealthCare Institutional Review Boards.

Interventions—The intervention group received 15 months of quality improvement support through exposure to a learning network, monthly interactive didactic webinars, quarterly face-to-face learning sessions and one of two practice improvement advisors who provided on-site coaching from 2011–2013. Based on a predefined driver diagram (Appendix), intervention practices were coached to perform rapid, small-scale tests of

change and to iteratively improve performance of high-risk clinical systems, as well as to embed simple measurement into routine work streams to guide improvement efforts.

Forming Practice Quality Improvement Teams—Each practice selected a team to lead the improvement effort that generally included frontline staff, a senior leader (e.g., medical director and/or practice manager), a clinical champion (e.g., an engaged physician who was a respected opinion leader), and a “day-to-day” champion (e.g., a nursing or administrative staff member) who helped to maintain momentum, convene and coordinate colleagues, and oversee implementation of change ideas.

Collaborative Learning Activities—Intervention practices participated in monthly webinars and six quarterly face-to-face learning sessions during the 15-month intervention period. The PROMISES team sequentially introduced key improvement concepts and facilitated discussion of practices’ process improvement efforts. Monthly webinars blended 30-minute didactic presentation with additional time for discussion. Quarterly face-to-face learning sessions were three-hour evening meetings including lectures from experts and presentations from intervention practices sharing their ongoing quality improvement efforts. These learning sessions covered a range of topics available on the PROMISES website.¹⁵

Improvement Advisors—Each practice was assigned a quality improvement advisor – an individual experienced in quality improvement methods and implementation – who visited the practices 1–2 times monthly to provide in-office coaching and follow-up on team progress. One-on-one conference calls were also provided as needed. Improvement advisors worked with practices to identify improvement opportunities and help plan small-scale tests of change using the Plan, Do, Study, Act (PDSA) Model for Improvement.¹⁶

Data Sources/Collection

Retrospective Chart Review—We conducted a retrospective review of up to 100 patient charts at each intervention site at baseline and post-intervention. To select charts, we identified an enriched sample of patients within a 6-month case ascertainment period, triggered by one of the following abnormal test results: creatinine (Cr) >1.8, potassium (K+) > 5.4, thyroid-stimulating hormone (TSH) >10, international normalization ratio (INR) >4, and prostate-specific antigen (PSA) >5. Additional abnormal laboratory values uncovered during the chart review and meeting specified threshold criteria during a 12-month look-back period were also included in data collection and assessment of documentation and follow-up. This targeted sample served to identify high-risk patients for critical test follow-up using lab triggers to identify an enriched sample of charts.¹⁷ We aimed to identify 20 charts for each test type. However, in practices with fewer than 20 abnormal results for a given test type (typically those with fewer providers), we adjusted by oversampling other critical laboratory results, where possible, to obtain a sample size of up to 100 charts per intervention site. Research staff worked with practices to implement the most practical method for identifying charts with abnormal results needed for review, which in most cases consisted of generating an electronic report from their electronic medical record or lab vendor. Due resource limitations and access to charts at control sites, chart review was performed only in intervention sites.

Study reviewers examined charts to determine if a) abnormal test results were present in the chart, b) the abnormality was noted by a responsible provider, c) there was documentation of an action or referral plan, d) documentation of patient notification of the abnormality, and e) evidence that the treatment or plan was completed for these abnormal results as well as other predefined high-risk results or findings (e.g. suspicious skin lesion, abnormal colonoscopy, imaging mass, suspicious breast mass, abnormal Pap smear). Reviewers recorded time elapsed between the date of the test and date of documentation/follow-up.

Trained research assistants abstracted data from patient charts under supervision of an experienced research nurse (CF) and general internist (GS). Inter-rater reliability of abstractors was assessed to ensure consistency in data abstraction ($\kappa=0.82$). Once identified, potential adverse findings detected during chart review were further assessed by the supervising general internist, who determined the severity of the potential risk, distinguishing findings as “serious potential risk” versus simply “potential risk” to patient safety. Any test result (meeting trigger criteria) where there was no documented evidence that the clinician was aware of the result or finding was treated as a potential patient safety risk. We defined serious potential safety risks as those events where we found potential or actual harm (i.e. if not treated, harm that would place the patient at risk of death or potential to cause persistent deterioration of life function). Any unresolved findings were also immediately fed back to the practices to take any needed clinical actions.

Chart reviews were designed to a) provide an objective measure of the impact of the intervention (complementing the more subjective staff and patient surveys), b) aid practices and improvement advisors in identifying areas for improvement that practices may not have previously recognized, and c) permit more standardized comparisons of variations among intervention sites.

Staff Surveys—We surveyed staff (including clinicians, managers, and administrative personnel) at intervention and control sites at baseline and post-intervention. The survey instrument was a 63-item questionnaire covering 11 domains: Access to Service and Care, Medication Management, Referral Management, Test Result Management, Malpractice Concerns, Patient-Focused Care, Quality and Risk Management, Practice Communication, Work Environment, Teamwork, and Practice Leadership. All items used 5-point scales: Likert (strongly agree to strongly disagree), frequency (always to never or daily to never), quality (excellent to poor), concern (extremely to not at all) and included a “not applicable” option. The study instrument and staff survey psychometric assessment has been described in detail elsewhere.¹³

Patient Surveys—We surveyed patients from intervention and control sites at baseline and post-intervention. The survey, based primarily on the CAHPS Patient-Centered Medical Home Survey^{18,19} was used to assess primary care patient experience. We selectively included or adapted additional questions and constructs from other validated instruments, such as the Patients Perceptions of Integrated Care (PPIC) Survey²⁰ and the Primary Care Assessment Survey (PCAS)^{21,22} and developed additional questions, resulting in a 34-item paper-based survey covering seven domains: Access, Communication, Coordination, Patient-centered Care, Office Flow, Trust, and, Quality of Primary Care Provider Care, plus

demographic and self-reported health status information. Items on the patient survey used 4-point frequency (always to never) or likelihood (definitely yes to definitely no) or 5-point Likert scales.

A professional research firm administered patient surveys by mail to patients identified from lists of patients seen during a 2–3 week period. As patients checked in for their appointments, we distributed letters printed on practice letterhead describing the survey and offering opportunity to opt out. Each waiting room also contained a poster displaying this information. Patients could decline having their contact information shared by notifying front-desk staff, writing their name on the informational letter and dropping it in a designated place in the medical office, or calling a designated phone number. Each practice then provided their list of remaining patients' names and addresses to the independent survey firm.

The survey firm randomly selected a subset of 150 patients to receive the survey by mail. An initial mailing included a cover letter, survey, pre-paid reply envelope to return the completed survey, and a \$5 incentive. Each survey was assigned a unique identification (ID) number. These IDs were used solely to identify surveys not yet returned. Three weeks after the initial mailing, non-responders were sent a reminder letter and, two weeks later, those who did not respond were sent a second copy of the survey.

To assess reliability of survey domains in the patient survey, we used Cronbach's alphas. We examined discriminant validity by comparing interscale correlations between domains to inter-item correlations for each domain. Cronbach's alphas ranged from 0.74 to 0.94 (mean 0.77) for 9 of the 11 domains; two three-item scales, medication management and access to services, did not meet standard reliability thresholds. Correlations between survey scales were lower than inter-item correlations for all comparisons, ranging from 0.05 to 0.69 (mean 0.36).

Intervention Fidelity Assessments—To assess intervention fidelity, our two PROMISES improvement advisors assigned scores to each practice assessing level of participation in PROMISES activities for projects related to the PROMISES 3+1 domains. Based on these scores, intervention sites were dichotomized into those with higher vs. lower intervention fidelity. This approach is described in detail elsewhere.²³

The study surveys, chart review instrument, and updated learning materials are all available online (www.brighamandwomens.org/pbrn/promises).

Statistical Analyses—We used the Student's t-test to compare means and the χ^2 and Fisher's exact test to examine differences in proportions among demographic characteristics between intervention and control practices. We used the SURVEYFREQ procedure in the SAS package to compare rates of documentation and follow-up of abnormal laboratory test results among intervention practices. Because we analyzed each laboratory test and individual patients could have multiple tests, the Rao-Scott method was used to adjust for clustering of patients within practices. We calculated incident rate ratios for potential and serious potential patient safety risks. Log-rank tests that accounted for censoring were used

to compare length of time between laboratory test order dates and dates that abnormal test results were documented/followed up among intervention practices. For both staff and patient surveys, after pairwise deletion of cases with missing values, we compared post-versus pre-test differences in perceptions overall and by safety domain using a difference-in-difference comparison. We reverse scored negatively-phrased questions such that a lower score was better for all items, regardless of original phrasing. While central tendency is an important indicator, a negative response indicates antithesis of the desired condition and may highlight substantial problems in the practice (such outliers raising concerns for malpractice risks). Accordingly, we examined percent of negative responses, scores of 4 or 5. We considered p-values < 0.05 statistically significant. All analyses were conducted using SAS (Version 9.3; Cary, North Carolina).

RESULTS

Practice Characteristics

Table 1 presents practice characteristics for intervention and control sites. Practices in the intervention group unexpectedly had significantly younger patients than those in the control group. Differences among all other practice demographics were non-significant.

Chart Review of Safety Risks

Across the 16 intervention sites, we reviewed 815 patient charts at baseline and 762 post-intervention, ranging from 17 to 100 charts per office in each period. Within these charts, we identified 1629 and 1530 abnormal lab tests pre and post-intervention, respectively. After the 15-month intervention, practices saw a reduction in rates of potential patient safety risks from 155 per 1000 patients with an abnormal lab value to 54 per 1000 [IRR 0.35 (95% CI 0.24 – 0.50)]. Serious patient safety risks (Table 2) decreased from 28 per 1000 patients with an abnormal lab value to 13 per 1000 [IRR 0.47 (95% CI 0.22–0.98)].

Change in clinician acknowledgement of abnormalities was the only process step that was non-significant in the before and after findings (0.2%, p=0.83; see Figure 1a). We found improvements in each stage of abnormal test follow-up: documentation of the abnormal results in chart (absolute improvement 1.4%, p=0.001), patient notification (5.8%, p<0.001), documentation of an action or treatment plan (6.1%, p<0.001), and evidence of a completed plan (8.6%, p<0.001) when comparing pre and post-intervention percentages. Decreases in length of time between date of laboratory test and documentation/follow-up of abnormal results were significant for time to document results in the chart (mean reduction of 0.41 days, p=0.02) and evidence that an action/treatment plan was completed (mean reduction of 19.4 days, p<0.001; see Figure 1b), but not significant for other outcomes.

Survey of Staff Perceptions and Attitudes: Intervention and Control Sites. (Table 3)

Across all 25 practices, 292 (61%) and 287 (60%) staff members responded to the pre- and post-intervention surveys, respectively. Staff generally reported mean scores close to 2 on a 5-point scale, where 1 is the best possible score. Patient-Focused Care and Access to Services and Care were ranked most highly (best) with a mean score of 1.61. Work environment and practice leadership had the lowest (worst) mean scores, 2.47 and 2.33,

respectively. Focusing on negative responses, staff were less negative regarding medication management (2%) and more negative regarding work environment and referral management at 20% and 18% respectively. Percent negative response was greater than 10% (the threshold typically targeted by high reliability organizations) for practice leadership, practice communication, teamwork, and malpractice concerns.²⁴

For all practices studied (intervention and control), although mean scores worsened slightly overall (by 0.02), they improved slightly (−0.01) for the three high-risk domains. Likewise, percentage of negative responses also increased slightly overall (0.42) but improved for the three high-risk domains (−1.67).

Intervention practices improved more compared to control practices overall (−0.06 for means and −1.56 for percent negative response) and for the 3 high-risk domains that were the focus of PROMISES' efforts (−0.06 for means and −1.36 for percent negative response). Scores for intervention practices improved relative to controls for 9 of 11 domains based on means, and 7 of 11 domains based on negative response, but few of these differences were significant. However, percent of negative responses regarding both test result management and teamwork did improve significantly in intervention practices compared to controls (−7.54, $p<0.001$ and −7.55, $p<0.05$, respectively).

Overall, practices with higher intervention fidelity were not found to have improved more than those with lower fidelity. Differences in mean scores were 0.15 overall and −0.17 for the 3 high-risk domains, an effect of 9%, i.e., comparing the difference to pre-test mean. Differences in percent of negative responses were 5.80 overall and −4.94 for the 3 high-risk domains ($p<0.10$), a 48% effect relative to baseline. Significant differences between practices based on intervention fidelity were few; however, practices with higher fidelity showed significantly greater improvement in areas of practice communication (−1.11 mean score, $p<0.001$ and −30.52 percent negative response, $p<0.001$), practice leadership (−.40 mean, $p<0.05$), and test result management (−6.35 percent negative response, $p<0.05$). In contrast, perceived “malpractice concerns” increased (worsened) among practices with higher intervention fidelity relative to lower intervention fidelity (1.08 mean score, $p<0.05$ and 26.42 percent negative response, $p<0.01$).

Patient Perceptions and Attitudes Survey (Table 4)

Across all 25 practices, 1767 (48%) and 1521 (42%) patients responded to the pre- and post-intervention surveys, respectively. Patient perceptions were mostly positive (mean 1.52 on a scale of 1 to 5 where 1 is best) and no improvement in the intervention vs. control practices was detected. According to patients, practices with high intervention fidelity improved slightly relative to those with low intervention fidelity (−0.06 improvement in score and −1.61 improvement in negative response) and significantly or marginally in some domains: communication (−0.11 mean, $p<0.05$ and −2.60 percent negative response, $p<0.10$), coordination (−0.09 percent negative response, $p<0.05$), patient-centered care (−0.08 mean, $p<0.05$), and office flow (−0.07 mean, $p<0.10$).

DISCUSSION

We performed a multi-method evaluation of an intervention aimed at helping small- and medium-sized primary care practices reduce malpractice and safety risk and found modest but significant improvements, particularly with regard to chart-based evidence of reliable follow-up on high-risk abnormal test results. Gains were also observed in several other targeted high-risk domains, including critical referrals and a number of measures of practice patient safety culture. Although significant improvement on many measures was not detected, given the numerous challenges faced by these and similarly busy primary care practices, the results are promising for showing evidence that some measurable improvements are possible.

The most noteworthy finding was dramatic reduction in number of “open loops” – situations where documented follow-up on tests and referrals was lacking.^{25–27} While it could be argued that such improvements merely represent better documentation rather than actual improved care or patient safety, from a medico-legal standpoint, better documentation has intrinsic value (i.e., “if it isn’t documented, it didn’t happen and can’t be defended”).²⁸ Further, our tracking of improvement teams’ effort suggests they devoted relatively little energy to improvement activities related to documentation per se; rather they concentrated on identifying failure modes in overall care processes related to (mainly) test result and referral management, targeting “tests of change” to redesign these processes and improve their reliability.

These encouraging chart review findings need to be tempered by the fact that, while the most objective measure in our study, they were unblinded and based on “before vs. after” comparisons rather than comparisons to randomized controls (since the study lacked resources for chart review at control sites). The enhanced quality of the care we documented may reflect improvement over time as, for example, as practices become more experienced and savvy in using their EMRs for tracking results and referrals, rather than as a result of PROMISES interventions. Arguing against this explanation, however, are findings from other studies of EMRs and malpractice risk that have failed to detect such improvements over similar timeframes.^{29–31}

Improvement initiatives as part of Patient Centered Medical Home (PCMH) projects and evaluations are being increasingly reported.^{32–36} However this is the first primary care collaborative intervention focused specifically on malpractice and safety risk. Practices implemented a total of 28 changes in their processes related to management of referrals (n=9), test results (n=7), medications (n=4), and communication (n=8), falling short of more widespread, activity our project had hoped to see and have been reported in other PCMH interventions.^{36,37} Nonetheless, the 16 intervention practices described benefiting considerably from their involvement with PROMISES, particularly from the hands-on coaching by improvement advisors.¹⁴ Given the distractions of day-to-day caring for a full schedule of patients as well as a host of competing and concurrent priorities and initiatives, coaching by improvement advisors’ and didactic and interactive activities were useful for keeping practice improvement teams engaged and focused on improvement work. (See Supplemental Table-Lessons)

Practices and the project faced many constraints, similar to those reported in other studies,^{32,38} which included severe time constraints, competing priorities, practice turnover (several practices had nearly 100% turnover over the course of the two-year collaborative), the relatively short duration of the intervention improvement period, as well as practice diversity (in terms of size, network affiliation, infrastructure, prior experience with quality improvement methods, staffing numbers and types, existing processes for referral and test result management, and EMR systems) limiting development and sharing of uniform interventions across practices.¹⁴³⁹

Our study had a number of limitations. As noted above, the chart reviews lacked comparison data from concurrent controls and thus improvements represented only before-and-after data. In contrast, we did have randomized control practice data for staff and patient surveys, which mostly showed modest but non-significant improvements. Although we randomly allocated practices to intervention vs. controls, intervention practices had significantly more younger patients than their counterpart control practices. Practice turnover also could have confounded our finding, although we had several high turnover practices in both control and intervention arms. Although we intended to have 16 practices in each study arm, several of the recruited practices dropped out (due to time constraints) prior to randomization. To maintain the full intervention learning potential, we randomized to keep the full number of intervention practices and allocated only 9 to serve as controls. Staff survey data were also limited in sample size and obviously represent subjective ratings of practice conditions and experiences. Because of both staff turnover, and the fact that different patients were surveyed for the baseline vs. post-intervention period (to preserve patient confidentiality), we were unable to make direct comparisons before and after the interventions with the same staff and patient survey participants. Other limitations include potential lack of generalizability, as study practices came from a single state and volunteered for the project, thus may not be representative of small-to-medium-sized practices. However, this potential “bias” is perhaps offset by the fact that some practices were “volunteered” by their parent network organization, meaning that some were more reluctant and/or less prepared to participate. On the other hand, some practices benefitted considerably by being part of one particular network that consistently and enthusiastically supported the project. Finally, chart reviews have potential for variability, although our inter-rater correlation achieved a substantial level of reliability.⁴⁰

In conclusion, a project targeting ambulatory malpractice and safety risks demonstrated that a package of improvement-oriented interventions was associated with a reduction in instances of unaddressed test results and referrals, and showed modest but non-significant improvement in other risks and patient safety-related communication and culture measures. Working with busy small- and medium-sized practices was extremely challenging due to practices’ time constraints, health information technology, staff turnover, and practice variations.¹⁴ Providing resources, particularly support of experienced improvement advisors, appeared to be valued and valuable to the practices. Whether investment in such resources can be justified and sustained based on its ability to improve care and prevent costly malpractice suits requires further testing, experience, and study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge Alice Bonner, Jason Boulanger, Namara Brede, Sue Butts-Dion, Caitlin Colling, Anne Huben Kearney, Michaela Kerrisey, Japneet Kwatra, Marykate O'Malley, Ann Louise Puopolo, Dale Rogoff, Pat Satterstrom, Diana Whitney and the staff of the participating Massachusetts PROMISES practices who helped in the activities of the PROMISES project. Preliminary results from this analysis were presented at the 37th Annual Meeting of the Society of General Internal Medicine, San Diego, CA on April 14th, 2014 and the New England Regional Meeting of the Society of General Internal Medicine, Boston, MA on March 7th, 2014.

Disclosure of Funding: This project was supported by grant number R18HS019508 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

References

1. Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. *Ann Intern Med.* 2006; 5:488–96.
2. Gandhi TK, Lee TH. Patient safety beyond the hospital. *N Engl J Med.* 2010; 363:1001–3. [PubMed: 20825311]
3. Wallace E, Lowry J, Smith SM, Fahey T. The epidemiology of malpractice claims in primary care: a systematic review. *BMJ Open.* 2013; 3
4. Wynia MK, Classen DC. Improving ambulatory patient safety: learning from the last decade, moving ahead in the next. *JAMA.* 2011; 306:2504–5. [PubMed: 22166611]
5. Zuccotti G, Sato L. Malpractice risk in ambulatory settings: an increasing and underrecognized problem. *JAMA.* 2011; 305:2464–5. [PubMed: 21673301]
6. Bishop TF, Ryan AM, Casalino LP. Paid malpractice claims for adverse events in inpatient and outpatient settings. *JAMA.* 2011; 305:2427–31. [PubMed: 21673294]
7. Schiff GD, Puopolo AL, Huben-Kearney A, et al. Primary care closed claims experience of Massachusetts malpractice insurers. *JAMA Intern Med.* 2013; 173:2063–8. [PubMed: 24081145]
8. Moore G, Showstack J. Primary care medicine in crisis: toward reconstruction and renewal. *Ann Intern Med.* 2003; 138:244–7. [PubMed: 12558374]
9. Plsek PE, Greenhalgh T. Complexity science: The challenge of complexity in health care. *BMJ.* 2001; 323:625–8. [PubMed: 11557716]
10. Phillips RL Jr, Bartholomew LA, Dovey SM, Fryer GE Jr, Miyoshi TJ, Green LA. Learning from malpractice claims about negligent, adverse events in primary care in the United States. *Qual Saf Health Care.* 2004; 13:121–6. [PubMed: 15069219]
11. Montano MF, Mehdi H, Nash DB. Annotated Bibliography Understanding Ambulatory Care Practices in the Context of Patient Safety and Quality Improvement. *American Journal of Medical Quality.* 2016; 31:29S–43S. [PubMed: 27535940]
12. Kachalia A, Mello MM. New directions in medical liability reform. *N Engl J Med.* 2011; 364:1564–72. [PubMed: 21506746]
13. Singer SJ, Reyes Nieva H, Brede N, et al. Evaluating ambulatory practice safety: the PROMISES project administrators and practice staff surveys. *Med Care.* 2015; 53:141–52. [PubMed: 25464161]
14. Schiff GD, Reyes Nieva H, Griswold P, et al. Addressing Ambulatory Safety and Malpractice: The Massachusetts PROMISES Project. *Health Services Research.* 2016; 51:2634–41. [PubMed: 27892623]
15. The PROMISES Project. 2016. Accessed 04/22/2016, 2016, at www.brighamandwomens.org/pbrn/promises

16. Weizman AV, Mosko J, Bollegala N, et al. Quality improvement primer series: launching a quality improvement initiative. *Clinical Gastroenterology and Hepatology*. 2016; 14:1067–71. [PubMed: 27211504]
17. Classen DC, Lloyd RC, Provost L, et al. Development and evaluation of the Institute for Healthcare Improvement global trigger tool. *J Patient Saf*. 2008; 4:169–77.
18. Davies E, Shaller D, Edgman-Levitan S, et al. Evaluating the use of a modified CAHPS survey to support improvements in patient-centred care: lessons from a quality improvement collaborative. *Health Expect*. 2008; 11:160–76. [PubMed: 18494960]
19. Scholle SH, Vuong O, Ding L, et al. Development of and field test results for the CAHPS PCMH Survey. *Med Care*. 2012; 50(Suppl):S2–10. [PubMed: 23064272]
20. Singer SJ, Friedberg MW, Kiang MV, Dunn T, Kuhn DM. Development and preliminary validation of the Patient Perceptions of Integrated Care survey. *Med Care Res Rev*. 2013; 70:143–64. [PubMed: 23161612]
21. Safran DG, Kosinski M, Tarlov AR, et al. The Primary Care Assessment Survey: tests of data quality and measurement performance. *Med Care*. 1998; 36:728–39. [PubMed: 9596063]
22. Stange KC, Etz RS, Gullett H, et al. Metrics for assessing improvements in primary health care. *Annu Rev Public Health*. 2014; 35:423–42. [PubMed: 24641561]
23. Kerrissey M, Satterstrom P, Leydon N, Schiff G, Singer S. Integrating: A managerial practice that enables implementation in fragmented health care environments. *Health care management review*. 2016
24. Singer SJ, Gaba DM, Geppert JJ, Sinaiko AD, Howard SK, Park KC. The culture of safety: results of an organization-wide survey in 15 California hospitals. *Qual Saf Health Care*. 2003; 12:112–8. [PubMed: 12679507]
25. Dalal AK, Pesterev BM, Eibensteiner K, Newmark LP, Samal L, Rothschild JM. Linking acknowledgement to action: closing the loop on non-urgent, clinically significant test results in the electronic health record. *J Am Med Inform Assoc*. 2015; 22:905–8. [PubMed: 25796594]
26. Schiff, G. Getting results: reliably communicating and acting on critical test results. Oakbrook Terrace, Ill.: Joint Commission Resources; 2006.
27. Sittig DF, Singh H. Defining health information technology-related errors: new developments since to err is human. *Arch Intern Med*. 2011; 171:1281–4. [PubMed: 21788544]
28. Mangalmurti SS, Murtagh L, Mello MM. Medical malpractice liability in the age of electronic health records. *N Engl J Med*. 2010; 363:2060–7. [PubMed: 21083393]
29. Fernandopulle R, Patel N. How the electronic health record did not measure up to the demands of our medical home practice. *Health Aff (Millwood)*. 2010; 29:622–8. [PubMed: 20368591]
30. Linder JA, Ma J, Bates DW, Middleton B, Stafford RS. Electronic health record use and the quality of ambulatory care in the United States. *Arch Intern Med*. 2007; 167:1400–5. [PubMed: 17620534]
31. Samal L, Wright A, Healey MJ, Linder JA, Bates DW. Meaningful use and quality of care. *JAMA Intern Med*. 2014; 174:997–8. [PubMed: 24733228]
32. Crabtree BF, Nutting PA, Miller WL, Stange KC, Stewart EE, Jaen CR. Summary of the National Demonstration Project and recommendations for the patient-centered medical home. *Ann Fam Med*. 2010; 8(Suppl 1):S80–90. S2. [PubMed: 20530397]
33. Friedberg MW, Rosenthal MB, Werner RM, Volpp KG, Schneider EC. Effects of a Medical Home and Shared Savings Intervention on Quality and Utilization of Care. *JAMA Intern Med*. 2015; 175:1362–8. [PubMed: 26030123]
34. Friedberg MW, Schneider EC, Rosenthal MB, Volpp KG, Werner RM. Association between participation in a multipayer medical home intervention and changes in quality, utilization, and costs of care. *JAMA*. 2014; 311:815–25. [PubMed: 24570245]
35. Kern LM, Edwards A, Kaushal R. The Patient-Centered Medical Home and Associations With Health Care Quality and Utilization: A 5-Year Cohort Study. *Ann Intern Med*. 2016; 164:395–405. [PubMed: 26882534]
36. Wagner EH, Coleman K, Reid RJ, Phillips K, Abrams MK, Sugarman JR. The changes involved in patient-centered medical home transformation. *Prim Care*. 2012; 39:241–59. [PubMed: 22608865]

37. Crabtree BF, Nutting PA, Miller WL, et al. Primary care practice transformation is hard work: insights from a 15-year developmental program of research. *Med Care*. 2011; 49(Suppl):S28–35. [PubMed: 20856145]
38. Reid RJ, Wagner EH. The Veterans Health Administration Patient Aligned Care Teams: lessons in primary care transformation. *J Gen Intern Med*. 2014; 29(Suppl 2):S552–4. [PubMed: 24715404]
39. Crabtree BF, Chase SM, Wise CG, et al. Evaluation of patient centered medical home practice transformation initiatives. *Med Care*. 2011; 49:10–6. [PubMed: 21079525]
40. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–74. [PubMed: 843571]

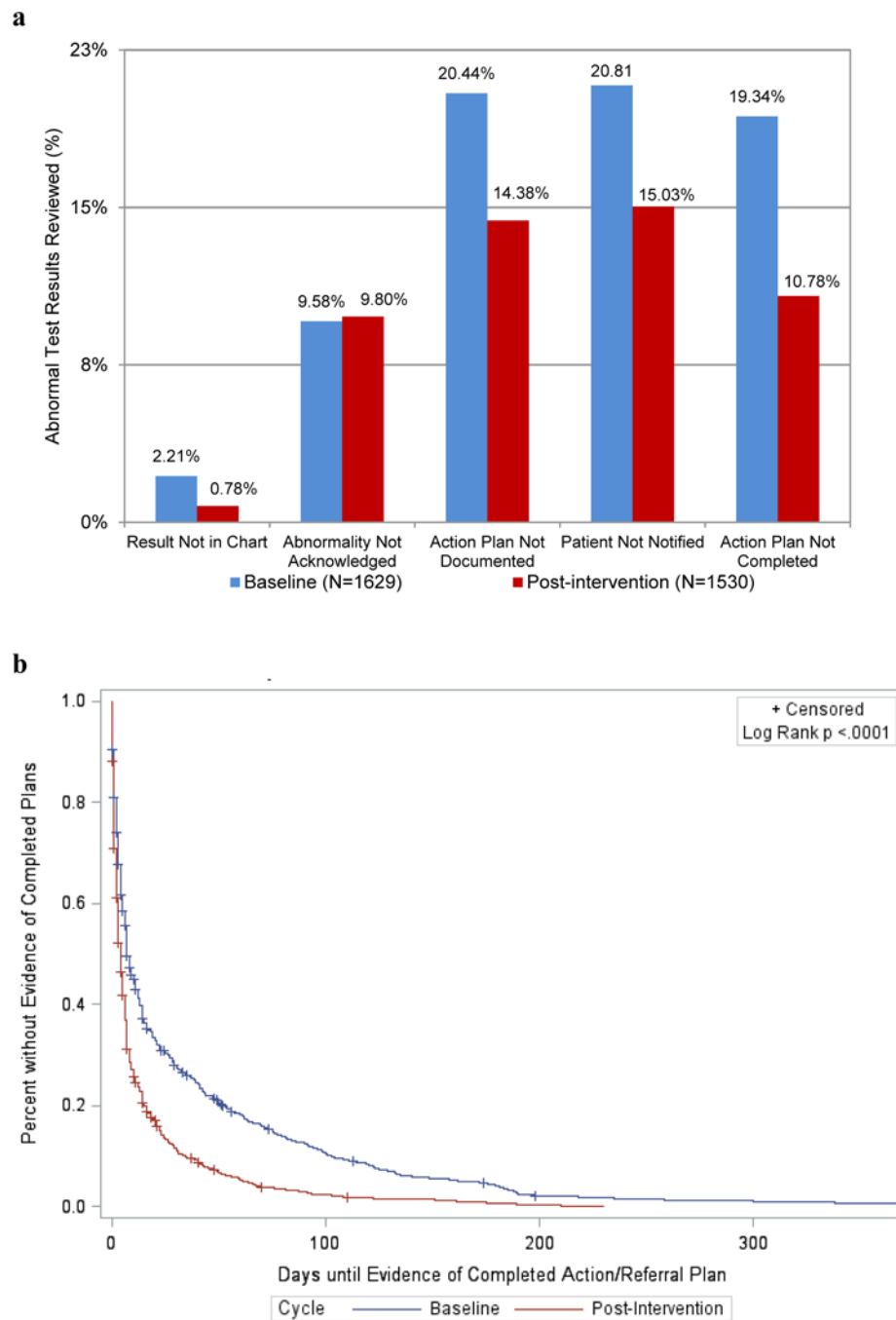


Figure 1. Abnormal test result documentation and follow-up among intervention practices. (a) Percentage of abnormal test results without appropriate documentation or follow-up at baseline and post-intervention. (b) Days until evidence of completed action plan or referral for abnormal test results at baseline and post-intervention.

Table 1

Primary Care Practice Characteristics by Study Arm

<i>Characteristic*</i>	<i>All Practices (N=25)</i>	<i>Intervention (N=16)</i>	<i>Control (N=9)</i>	<i>p-value</i>
Number of staff in practice by role, mean				0.999
Physicians (MD or DO)	5	4	6	
Physician Extenders (PA, NP, CNS, APN)	1	1	2	
Nursing (RN, LVN, LPN)	1	1	1	
Clinical Support (MA)	5	4	6	
Management	1	1	2	
Administrative	5	4	7	
Number of staff years at practice, median				0.186
2 month to less than 1 year	2	2	2	
1 year to less than 3 years	6	3	5	
3 years to less than 6 years	6	6	6	
6 years to less than 11 years	2	4	4	
11 years or more	2	0	7	
Number of staff hours worked per week, median				0.228
17 to 24 hours per week	0	0	3	
25 to 32 hours per week	2	2	0	
33 to 40 hours per week	13	11	17	
41 hours per week or more	3	2	4	
Number of patients per FTE physicians, mean	1462	1403	1580	0.658
Percentage of patients by age group, mean				<.001
19–39 years of age	26	31	23	
40 – 64 years of age	46	47	46	
65 years of age	28	22	31	
Practice Uses EMR System, [N(%)]				0.444
Yes	24 (96)	15 (94)	9 (100)	
No	1 (4)	1 (6)	0 (0)	
Practice Uses Electronic Prescribing Module, [N(%)]				0.444
Yes	24 (96)	15 (94)	9 (100)	
No	1 (4)	1 (6)	0 (0)	
Practice Holds Regular Staff Meetings, [N(%)]				0.444
Yes	24 (96)	15 (94)	9 (100)	
No	1 (4)	1 (6)	0 (0)	
Practice Regularly Reviews and Discusses Safety and Reliability Issues with Staff, [N(%)]				0.629
Yes	18 (72)	11 (69)	7 (78)	
No	7 (28)	5 (31)	2 (22)	

* Number of staff (by profession) and patients (by age) reported are mean values across practices; staff years and staff hours reported are median values across practices; CNS, certified nurse specialist; DO, doctor of osteopathy; FTE, full-time equivalent (40 hours per week); MD, medical doctor; NP, nurse practitioner; PA, physician's assistant; PCP, primary care provider

Table 2

List of all serious potential patient safety risks identified by chart review

<i>Practice</i>	<i>Baseline</i>	<i>Post-Intervention</i>
1	Two cases of failure to follow up abnormal CT and lab (high PSA recommended urology but no referral found)	None
2	Scalp nodule suggestive for cancer; no documented referral	None
3	Outstanding Derm referral for suspicious skin lesion	Elevated Cr 2.1 (CrCl 35.85): Pt on Metformin. Pt also on Glyburide. Recommended Glipizid.
4	Pt informed of high PSA (doubled in a year). No subsequent encounters or tests documented	None
5	Abnormal TSH not followed up; Pulmonary nodule not followed-up	Pt's CrCl is 41.12 and prescribed Allopurinol 300mg. The maximum safe dosage is 200mg.
6	High PSA referral in 1mo recommended, still outstanding after 10 months	None
7	Abnormal TSH and PSA; no follow-up	1. PCP acknowledged lung nodules found on CT and need for follow-up. No evidence of follow-up in chart. 2. Chronic lung nodule, 4 mm, had CT scan with changes. No referral in chart. 3. PCP aware and notified pt of high TSH and importance of treatment of hypothyroidism but needed to continue to be followed up. No evidence of treatment after 2 months.
8	High PSA referral outstanding	Urology referral done and +CA but no documentation if pt and PCP aware of result
9	Liver lesion no follow-up recommended by radiology	Elevated K 5.6: No evidence this was acknowledged, pt notified, or followed up. Pt has had no new K since.
10	Thyroid nodule/cyst; no recommended repeat ultrasound	Pt has high Cr (1.9) last checked on 9/3/2013 and has been on renal contraindicated Metformin.
11	None	1. Elevated Cr of 2.02: Patient's CrCl was 42.47 and currently on 300mg of Allopurinol to take daily. 2. Elevated Cr 2.12: Patient is on Warfarin and is prescribed Ibuprofen.
12	Kidney cyst on ultrasound not followed up; PSA not followed up	None
13	Note that pt had underactive thyroid, recommended TSH retest. No follow up 1 year later.	None
14	Likely basal cell cancer –outstanding referral	None
15	Derm CA referral outstanding; Pulmonary nodule on CT missed; follow-up on high TSH missing	None
16	New abnormal TSH not followed-up	None

* CA, cancer; Cr, creatinine; CrCl, creatinine clearance; CT, computerized tomography scan; Derm, dermatology; K, potassium; PCP, primary care physician; Pt, patient; PSA, prostate-specific antigen; TSH, thyroid stimulating hormone

Table 3

Comparison of Staff Perceptions at Intervention and Control Practices

Domain Name	Pre (n)	Post (n)	Mean score				Percent (%) negative response			
			Post		Pre-post difference		Post		Pre-post difference	
			All practices	Higher vs. Lower Intervention Fidelity	All practices	Higher vs. Lower Intervention Fidelity	All practices	Higher vs. Lower Intervention Fidelity	All practices	Higher vs. Lower Intervention Fidelity
Access to Service and Care	276	271	1.61	0.07	-0.12	-0.09	3.57	1.27	-2.32	-3.16
Medication Management	269	264	1.62	0.04	-0.05	0.09	2.27	-0.58	1.39	1.77
Referral Management	269	273	2.30	-0.13	-0.10	-0.20	18.41	-5.23	0.64	-8.97
Test Result Management	252	257	1.72	0.02	-0.18	-0.04	6.18	0.29	-7.54***	-6.35*
Malpractice Concerns	249	247	1.86	-0.07	0.198	1.08*	10.53	-1.04	5.12	26.42**
Patient-Focused Care	256	258	1.61	0.04	-0.15 [^]	0.01	4.03	-0.28	-0.94	1.20
Quality and Risk Management	257	251	2.01	0.04	-0.13	0.38	7.57	-0.80	-5.31	14.75
Practice Communication	260	251	2.15	0.15	0.01	-1.11***	12.35	2.95	-1.78	-30.52***
Work Environment	261	247	2.47	0.11	-0.11	0.26	20.29	3.91 [^]	-3.16	4.56
Teamwork	262	244	2.21	0.24*	-0.08	-0.31	11.61	5.21*	-7.55*	-7.22
Practice Leadership	231	216	2.33	0.20*	-0.05	-0.40*	14.24	2.66	1.53	1.41
Overall Average			1.76	0.02	-0.06	0.15	8.85	0.42	-1.56	5.80
<i>Simple effect (difference/pre mean)</i>				1%	-3%	9%		5%	-19%	69%
<i>Average Med. Referral, Test Results</i>			1.78	-0.01	-0.06	-0.17	8.59	-1.67	-1.36	-4.94[^]
<i>Simple effect (difference/pre mean)</i>				-1%	-3%	-9%		-16%	-13%	48%

Notes:

[^] p-value <.1;

* p-value<.05,

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

** p-value<.01,

p<.001; p-values <.05 considered statistically significant.

Items are scored on a 5-point scale ranging from 1 (best) to 5 (worst); lower score and negative difference are better.

Table 4

Comparison of Patient Perceptions at Intervention and Control Practices

Domain Name	Pre (n)	Post (n)	Mean score				Percent (%) negative response			
			Pre-post difference		Post		Pre-post difference		Post	
			All practices	Intervention vs Control	All practices	Higher vs. Lower Intervention Fidelity	All practices	Intervention vs Control	All practices	Higher vs. Lower Intervention Fidelity
Access	1559	1354	1.60	-0.05	-0.06	-0.06	4.06	-1.83*	-0.59	0.46
Communication	1558	1365	1.76	0.02	-0.05	-0.11*	9.26	0.21	-0.70	-2.60 [^]
Coordination	1558	1363	1.76	-0.06	-0.04	-0.09*	9.09	-0.79	-0.38	-5.19**
Patient-Centered Care	1557	1363	1.39	-0.02	0.003	-0.08*	0.93	-0.32	-0.61	-1.00
Office Flow	1560	1357	1.65	-0.02	-0.05	-0.07 [^]	4.06	-0.44	-0.59	-0.17
Trust	1545	1360	1.59	0.00	0.04	—	3.38	-0.18	-0.21	—
Quality of PCP Care	1549	1365	1.50	0.02	0.03	—	2.56	-0.19	0.28	—
Overall Average			1.52	-0.02	-0.03	-0.06	4.52	-0.45	-0.41	-1.61
Simple effect (difference/pre mean)				-1%	-2%	-4%		-9%	-8%	-32%

Notes:

[^]p-value <1;

* p-value<.05,

** p-value<.01,

p<.001; p-values <.05 considered statistically significant.

Items are scored on a 5-point scale ranging from 1 (best) to 5 (worst); lower score and negative difference are better. For Trust and Quality of PCP care patient survey domains, improvement advisors were unable to rate the practices based on the type of contact they had with their practices in the course of their improvement work.