

# Project RedDE: Cluster Randomized Trial to Reduce Missed or Delayed Abnormal Laboratory Value Actions

Michael L. Rinke, MD, PhD\*<sup>†</sup>; David G. Bundy, MD, MPH<sup>†</sup>; Christoph U. Lehmann MD<sup>‡</sup>; Moonseong Heo, PhD<sup>§</sup>; Jason S. Adelman MD, MS<sup>¶</sup>; Amanda Norton MSW<sup>||</sup>; Hardeep Singh, MD, MPH\*\*

## ABSTRACT

**Introduction:** Failure of timely abnormal laboratory result follow-up is relatively common and may lead to harm. This study hypothesized that a quality improvement collaborative (QIC) could reduce the frequency of missed or delayed action on abnormal laboratory values. **Methods:** A national cohort of pediatric practices was cluster-randomized to sequentially receive a QIC intervention: video conferences, transparent data sharing, a “focus on failures,” QI coaching, and tools to help reduce missed or delayed action on abnormal laboratory values. Practices recorded the percentage of patients with 5 specific abnormal laboratory values who received an appropriate provider action (control), and then, during an 8-month intervention phase, implemented QI strategies to reduce errors (intervention). Subsequently, practices collected data on laboratory errors while working to reduce unrelated second (sustain phase), and third (maintenance phase) errors. Generalized mixed-effects regression models compared the mean percentage of patients with appropriate actions. **Results:** We randomized 43 practices, of which 31 were included in analyses. Control and intervention phases included 1,357 and 1,426 patients with abnormal laboratory values, respectively. The mean percentage of patients who received appropriate actions did not change comparing control and intervention phases [risk difference (RD) 1%; 95% CI -1%, 3%]. In *post-hoc* analyses, practices significantly improved comparing control to sustain (RD 3%; 95% CI 0.3%, 6%) and maintenance phases (RD 6%; 95% CI 3%, 9%). **Conclusion:** Implementation of a QIC did not reduce the frequency of abnormal laboratory errors in the initial 8-month intervention phase. A significant reduction was appreciated comparing sustain and maintenance phases (months 9–24) to the control phase. (*Pediatr Qual Saf* 2019;4:e218; doi: 10.1097/pq9.000000000000218; Published online September 30, 2019).

## INTRODUCTION

Failure to respond to abnormal laboratory results can lead to serious harm,<sup>1–3</sup> and is a frequent cause of malpractice lawsuits and payouts.<sup>4,5</sup> Forty percent of ambulatory primary care visits include laboratory testing,<sup>6</sup> and in one study, 83% of physicians reported at least 1 delay in reviewing laboratory results during the previous 2 months.<sup>7</sup> Forty percent of physicians reported missing abnormal results despite a



highly computerized health system.<sup>1</sup> In ambulatory patients, failure to respond appropriately to tests can occur with the majority of laboratory results.<sup>3,5,8</sup> Often physicians rely on subsequent patient visits to identify abnormal results,<sup>9</sup> suggesting that if a patient fails to return for a subsequent visit, abnormal laboratory results would never be noticed. This lack of response can lead to error and patient harm.<sup>2,10,11</sup> While laboratory follow-up is a recognized problem

From the \*Department of Pediatrics, The Children's Hospital at Montefiore and the Albert Einstein College of Medicine, Bronx, N.Y.; †Department of Pediatrics, Medical University of South Carolina, Charleston, S.C.; ‡Departments of Biomedical Informatics and Pediatrics, Vanderbilt University Medical Center, Nashville, Tenn.; §Department of Public Health Sciences, and School of

Mathematical Sciences, Clemson University, Clemson, S.C.; ¶Department of Internal Medicine, Columbia University College of Physicians and Surgeons, New York, N.Y.; || A. Mandatory, Inc., Ithaca, N.Y.; and \*\* Center for Innovations in Quality, Effectiveness and Safety, Michael E. DeBakey Veterans Affairs Medical Center, Department of Medicine, Baylor College of Medicine, Center of Innovation, Houston, Tex.

Copyright © 2019 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

This work and the authors, M.L.R., H.S., D.G.B., and M.H., are supported by the Agency for Healthcare Research and Quality (HS0203608-01). H.S. is partially supported by the Houston VA HSR&D Center for Innovations in Quality Effectiveness and Safety (CIN 13-413).

To cite: Rinke ML, Bundy DG, Lehmann CU, Heo M, Adelman JS, Norton A, Singh H. Project RedDE: A Cluster Randomized Trial to Reduce Missed or Delayed Action on Abnormal Pediatric Laboratory Values. *Pediatr Qual Saf* 2019;5:e218.

Supplemental digital content is available for this article. Clickable URL citations appear in the text.

Clinical Trial Registration: Reducing Diagnostic Errors in Primary Care Pediatrics (Project RedDE), NCT02798354, clinicaltrials.gov

Received for publication February 4, 2019; Accepted August 27, 2019.

Published online September 30, 2019.

DOI: 10.1097/pq9.000000000000218

\*Corresponding author. Address: Michael L. Rinke, MD, PhD, Department of Pediatrics, The Children's Hospital at Montefiore and the Albert Einstein College of Medicine, 3411 Wayne Avenue, Room 801, Bronx, NY  
PH: 718-741-2524; Fax: 718-920-4351  
E-mail: mrinke@montefiore.org

in adult patients, studies on children are lacking, and even fewer studies have worked to improve laboratory follow-up in this vulnerable population.

Our prior work demonstrated that reducing high-frequency/sub-acute errors, such as missed or delayed action on abnormal laboratory values, is of strong interest to primary care pediatricians.<sup>12</sup> Despite the high interest, we reported that only one-third of pediatricians were involved in efforts to reduce missed or delayed action on abnormal laboratory values in their practices.<sup>12</sup> Data from pediatric primary care practices found 11% of patients with 5 specific abnormal laboratory values did not have appropriate and timely action documented.<sup>13</sup> It is unclear what system-level strategies can be used to improve testing processes and reduce the harm from missed or delayed action on abnormal laboratory values in children.

This study aimed to determine whether a quality improvement collaborative (QIC) intervention that offered multiple improvement strategies could reduce the frequency of missed or delayed action on abnormal laboratory values in pediatric practices and sustain those reductions. We tested this hypothesis via a prospective, stepped wedge cluster randomized controlled trial in a national cohort of pediatric primary care clinics.

## METHODS

As described previously,<sup>13,14</sup> Project RedDE (*Reducing Diagnostic Errors in Pediatric Primary Care*) aimed to reduce 3 related diagnostic errors in primary care pediatric practices in collaboration with the American Academy of Pediatrics' (AAP) Quality Improvement Innovation Networks (QuIIN). QuIIN works to "improve the quality and value of care and outcomes for children and families" via quality improvement networks. Missed or delayed action on abnormal laboratory results was 1 of the 3 errors addressed.

### Randomization

Overall, 43 practices were recruited in 2 waves and randomized via computer random number generator in a non-blinded fashion. We included 31 practices in the final analysis. In March 2015, 34 pediatric "Wave 1" practices were randomized to 1 of 3 groups with multivariate matching before randomization<sup>15</sup> based on university affiliation, the presence of self-reported prior efforts to reduce the target errors, and total annual visits per pediatric practitioner equivalents. Nine practices dropped out after randomization but before submitting data; all were unable to collect necessary data. Of the remaining 25 Wave 1 practices, 24 submitted complete project data through September 2017; 1 practice dropped out after providing data for 8 months after their lead physician left the practice. We included this practice's control data in analyses. Nine additional "Wave 2" practices were recruited in December of 2015 to increase the sample size of the cohort. Of these, 2 practices dropped out after randomization but before

submitting data due to data burden; 2 other practices from a single care network merged into 1 team to boost their effective practice sample size. The resulting 6 Wave 2 teams participated alongside the 25 Wave 1 teams; Wave 2 teams participated in 2 action periods, while Wave 1 teams participated in 3 (Figs. 1, 2).

### Study Design

In July 2015, each of the 3 groups was randomized to collect retrospective baseline data (February–June 2015) on 1 of 3 errors: missed or delayed action on abnormal laboratory values, missed elevated blood pressure, or missed recognition or diagnosis of adolescent depression.<sup>13</sup> After an additional month of prospective baseline data (September 2015, which is combined for analyses with the 5 months of retrospective baseline data, and collected to ensure prospective data collection was similar to retrospective data collection), the groups began an 8-month QI action period to reduce their assigned first error. Concurrently (September 2015–May 2016), each group was assigned to collect control data on a second error. In a prospective, stepped-wedge fashion, during a second action period, (June 2016–January 2017) each group started work on reducing a second error, sustain the improvement on their first error, and collect control data for the third error. Finally, in February 2017, each group started working to reduce their third error, sustain improvements on their second error, and maintain improvements on their first error while receiving reduced QIC feedback and attention on the first error (Fig. 1).

Thus, each group had a "control phase" of laboratory error data collection without any QIC focus on reducing them, and all but one Wave 2 group had an "intervention phase" where clinic teams specifically worked to reduce missed or delayed action on laboratory errors. Two groups had a "sustain phase" targeting a second, unrelated error while sustaining improvements on laboratory errors; 1 group had a "maintenance phase" where they targeted 2 unrelated errors and maintained improvements on laboratory errors.

### Intervention

The primary intervention, a QIC, is an organized, multifaceted approach to QI with (1) a target for improvement with large variation in current practice; (2) clinical and QI experts sharing best practice knowledge; (3) multidisciplinary teams from multiple sites willing to improve care; (4) a model for improvement with measurable outcomes for improvement, data feedback to teams, and small tests of change; and (5) a series of structured activities to advance improvement and share experiences of participating teams.<sup>16</sup> In Project RedDE, each practice identified a 3-person QI team consisting of a physician, a nurse, and another professional (eg, administrator, business associate, front desk staff, etc.). After completing baseline data collection and entry into an AAP-run web-based portal, teams participated in a 2-day interactive video conference on QI

Group	February 2015-June 2015 & September 2015	Sept. 2015	Action Period I: October 2015-May 2016	June 2016	Action Period II: June 2016-January 2017	Feb. 2017	Action Period III: February 2017-September 2017
<b>Wave 1</b>							
Group 1	Control Phase	Virtual Learning Session 1	Intervention Phase	Virtual Learning Session 2	Sustain Phase*	Virtual Learning Session 3	Maintenance Phase
Group 2	*		Control Phase		Intervention Phase		Sustain Phase
Group 3	*		*		Control Phase		Intervention Phase
<b>Wave 2**</b>			November 2015-March 2016				
Group 1			*		*		Control Phase
Group 2			Control Phase		Intervention Phase		Sustain Phase
Group 3			*		Control Phase		Intervention Phase

**Fig. 1.** Project RedDE Timeline for Missed or Delayed Action on Abnormal Laboratory Values. \*Practices were involved in Project RedDE during this time but working exclusively on the 2 non-laboratory errors. Practices in Groups 2 and 3 had already worked to reduce 1 or 2 other diagnostic errors, respectively, before beginning to work on laboratory errors. †During the Sustain and Maintenance Phases, practices began working to reduce a second and third diagnostic error respectively. \*\*Wave 2 practices integrated alongside Wave 1 practices, intervening first on Wave 1’s second diagnostic error. These practices never intervened on a third diagnostic error.

methodology and laboratory error-specific content. They then received rapid, transparent data feedback on performance monthly with aggregate collaborative benchmarking in the form of run charts emailed from the project leadership team and coaches, participated in monthly hour-long video conferences and completed monthly mini-root cause analyses. These mini-root cause analyses asked clinics to identify a patient with a laboratory error in their clinic and then to examine 15 standardized patient and systems factors that could have contributed to this error.<sup>17,18</sup> Run charts were presented monthly for all clinics and discussed as part of each monthly video call. We present the end of collaborative summary results as small multiples p-charts by group and wave in the Supplemental Digital Content at <http://links.lww.com/PQ9/A136> for Figure. Each practice had a QI coach provided by the project, and each group had an interactive email listserv and group-specific website with project resources. Day-long video conferences were conducted every 8 months as practices transitioned to new errors (Fig. 1) and monthly video conferences provided synchronized data feedback in the form of run charts, group discussion, and didactic education. When working on their third error, conferences presented run charts from both teams’ second and third errors monthly while run charts from their first error, including all phases of the project, were only presented quarterly. Practices could always access all of their run charts independently from a web-based data repository and were encouraged by coaches and the leadership team to evaluate and provide insight into their data.

The project leadership developed a guidance “toolkit” for each process step associated with laboratory review: (1) test results returning to the clinic, (2) provider viewing test results, (3) recognizing abnormal results, (4) notifying families or patients of abnormal results, (5) taking action on abnormal results, and (6) documenting action

taken. The foundation of this toolkit was the Agency for Healthcare Research and Quality’s “Improving Your Laboratory Testing Process” Toolkit<sup>6</sup> as well as Safety Assurance Factors for Electronic Health Record (EHR) Resilience Guide for test results reporting.<sup>19</sup> Standard workflows were encouraged, especially around designating responsible providers for laboratory value follow-up on weekends and holidays, and timely review of EHR laboratory value inboxes. Each team utilized and modified resources most relevant to supporting their internal processes. We maintained all resources on the Project RedDE website and made them available to the public following the project’s conclusion.<sup>20</sup>

**Measures**

We developed pragmatic measures and efficient data collection methods for busy clinicians. The primary outcome was the proportion of patients with any of 5 specific abnormal laboratory values who had appropriate action documented without delay per 100 patients (Table 1). These 5 sub-acute results (microcytic anemia, elevated lead level, sexually transmitted disease, streptococcal pharyngitis on culture only, possible hypo- or hyperthyroid) were selected because each test is frequently ordered in primary care, and unrecognized or untreated results can lead to harm.<sup>21-26</sup> Definitions of “appropriate actions” and delays, described in Table 1 and created by discussions with the QIC expert group, literature reviews, and local pilot testing, were necessarily broad because more detailed research-team led chart review was beyond the scope of this study.

A secondary outcome measure included the number of patients with abnormal laboratory values as defined above, where the provider documented that the result was “abnormal” or provided a corresponding diagnosis (eg, anemia, syphilis, etc.), and/or documented the appropriate

action as above per 100 patients. This measure captures the number of times a pediatrician may recognize the diagnosis, but either knowingly or unknowingly does not take a recommended action.

Each practice examined the first 10 patients each month who met inclusion criteria (Table 1) with one of these abnormal laboratory values and entered error data and demographics including age, sex, and insurance status, into a web-based portal. Charts were reviewed 30 days after the laboratory value resulted in allowing time for the clinician to take appropriate actions. Discussions of error definitions facilitated chart review training during webinars, informational slides, email list serves, and continuous research team availability.

As a process measure, practices evaluated EHR inboxes of 10 providers each month and documented the percent of providers with no unread or unacknowledged laboratory results for >72 hours. Once they reached >90% compliance for 2 consecutive months or their practice's initial 8-month intervention phase ended, they stopped collecting this measure. Participating centers collected all other measures for the entire period following a group's entry into the control phase for that error.

### Statistical Analysis

Using patients as the units of analysis, we compared the primary outcome of the mean number of patients with 1 of the 5 specific abnormal laboratory values with an appropriate provider action taken per 100 patients between the intervention and control phases. We present the primary outcome effect measures as model-based estimates of risk differences (RD) comparing control versus intervention phases. Generalized mixed-effects regression models with identity link adjusted for age, sex, insurance status, wave, and laboratory value were applied with month-specific, and practice-specific intercepts considered random. We revised and finalized our power analysis based on error rates estimated from the Wave 1 baseline data.<sup>13</sup> The minimally detectable effect size in terms of RD with >80% power at a two-sided significance level = 0.05, considering correlations of outcomes across months within practices

= 0.05 and correlations of outcomes across charts within periods = 0.5, was determined between control and intervention phases as  $RD \geq 4.2\%$ . Wave 2 practice recruitment allowed for us to maintain adequate power despite clinic attrition. We excluded patients with incomplete demographic data from the final analysis.

We additionally examined our primary and secondary outcomes using statistical process control p-charts, with Nelson Rules<sup>27</sup> signifying changes from baseline performance. On these charts, the intervention's initiation was adjusted, so each group began the intervention on "month 1." We created aggregated p charts for the primary and secondary outcomes, and small multiple p charts to identify trends across the 3 groups and variation between specific clinics.

A priori, similar mixed-effects regression models examined the secondary outcome, as well as differences between (1) intervention and sustain phases and (2) sustain and maintenance phases (Fig. 1). These analyses investigated whether practices could sustain and/or maintain improvements while working on other errors and with less focus from the QIC. We also examined different effects by laboratory value. Process measure data, represented by time to 2 consecutive months with >90% compliance, were compared between collaborative groups using Kaplan-Meier Analysis and log-rank test to test whether the process performance was equal across all 3 groups.

*Post-hoc* analyses, identified after results from primary analyses were known, were conducted comparing (1) the primary outcome between the control and sustain phases for all groups, (2) the control and maintenance phases for all groups, and (3) the control and intervention phases for just the first group randomized to work on this error. The former analyses were conducted to examine if improvement on these errors required >8 months to complete in a QIC intervention. The latter analysis may suggest if a lack of improvement in the primary outcome was due to ascertainment bias. Practices not randomized to reduce abnormal laboratory value errors were aware that they would ultimately be working on these errors and may have begun improvement work while still in the control

**Table 1. Inclusion Criteria for Abnormal Laboratory Values and Definitions of Appropriate Actions and Delays**

Laboratory Diagnosis	Abnormal Value	Ages	Appropriate Action(s)	Delay Definition* (days)
Microcytic anemia	Hemoglobin <11 g/dL and mean corpuscular volume <75 fL	1 and 2 year olds	Starting iron supplementation, sending iron studies, or family conversation on increasing dietary iron intake	30
Elevated lead level	Blood lead >5 µg/dL	1, 2, and 3 year olds	Family conversation on lead remediation or plan to retest for elevated lead levels	30
Sexually transmitted disease	Positive <i>Neisseria gonorrhoeae</i> , <i>Chlamydia trachomatis</i> , <i>Treponema pallidum</i> , or human immunodeficiency virus	>10 years old	Antibiotics started or referral to infectious disease specialist	7
Streptococcal pharyngitis on culture only	Positive group A streptococcal throat culture with negative rapid test	>1 year old	Antibiotics started or family conversation about positive test	7
Possible hypo- or hyperthyroid	Thyroid stimulating hormone (TSH) <0.5 or >4.5 µIU/mL	>1 year old	Plan to repeat TSH test or referral to endocrinologist	7

\*A delay was defined as no appropriate action within the listed number of days from the date the laboratory value resulted.

**Table 2. Demographics of Included Practices at Baseline: N (%)**

Characteristic	Group 1	Group 2	Group 3	All
No. practices	10	10	11	31
University affiliation	5 (50)	6 (60)	7 (64)	18 (58)
Previously worked "a lot" on 1 of the 3 errors of interest: percent yes	5 (50)	5 (50)	5 (45)	15 (48)
Total annual visits per full time physician or physician extender equivalents: mean (SD)	3,399 (1,671)	4,549 (2,875)	3,172 (1,274)	3,689 (2,057)
Patient demographics percentage: mean (SD)				
White, non-Hispanic/Latino	28 (10)	38 (21)	43 (22)	37 (19)
Hispanic/Latino origin	31 (18)	17 (9)	20 (20)	22 (17)
Black/African American	31 (22)	30 (24)	32 (19)	31 (21)
Asian	5 (4)	3 (3)	3 (3)	4 (3)
Native Hawaiian/other Pacific Islander	1 (3)	0.1 (0.3)	—	0.4 (2)
American Indian/Alaska Native	0.5 (1)	0.3 (0.5)	—	0.3 (1)
Other	4 (6)	12 (13)*	2 (5)	6 (9)
Team's quality improvement skill and knowledge at project enrollment (%)				
Very knowledgeable	1 (10)	1 (10)	—	2 (7)
Knowledgeable	5 (50)	7 (70)	6 (55)	18 (58)
Somewhat knowledgeable	4 (40)	1 (10)	5 (45)	10 (32)
Not knowledgeable	—	1 (10)	—	1 (3)

All practices used an EHR.

\*Two practices in this group had a large Somali population.

phase given their non-blinded enrollment, data collection during control phases, and awareness of the eventual need to reduce these errors. We completed all data analyses with SAS v9.4. Statistical process control p-charts were created with Minitab 17 Statistical Software. The AAP's and the Albert Einstein College of Medicine's Institutional Review Boards approved this study.

## RESULTS

We present the demographics of the 31 practices included in the primary analysis in Table 2. Data on 1,357 patients were available for control and on 1,426 patients for the intervention phase (Fig. 2). Due to missing insurance data or missing laboratory test data, we excluded 193 patients (7%). Patient demographics and specific laboratory tests included are presented in Table 3.

The model-based estimate mean percentage of patients with one of the specific abnormal laboratory values who received an appropriate action was unchanged from 93.0% in the control phase to 94.1% in the intervention phase (RD 1.1%; 95% CI -1.0%, 3.1%;  $P = 0.302$ ). We observed similar results for other a priori analyses (Table 4). Specifically, the secondary outcome which allowed for providers who may have documented an abnormal laboratory result but either intentionally or unintentionally did not act without a delay was not different comparing the intervention and control groups (RD 0.1%; 95% CI -1.8%, 1.9%;  $P = 0.922$ ). Although stabilities at the individual practice level might not have been achieved (see Supplemental Digital Content at <http://links.lww.com/PQ9/A136> for Figure), overall significant slopes were not identified when looking at data within individual phases (control, intervention, sustain, and maintenance phases), which suggest stability in each analysis group.

In post-hoc analyses comparing sustain and control phases as well as maintenance and control phases, practices significantly improved (RD 3.0%, 95% CI 0.3%, 5.7%;  $P = 0.03$ , and RD 5.9%, 95% CI: 2.5%, 9.2%,  $P = 0.001$ , respectively). When examining data from only the first group of 10 practices targeting these errors without the potential for ascertainment bias, practices significantly improved during intervention phase from 85.6% to 91.0% (RD 5.4%, 95% CI 1.6%, 9.2%;  $P = 0.006$ ). (Table 4)

Figure 3 presents the primary and secondary outcomes' p-charts, which align all groups' first intervention month as month 1. A significant shift occurred after month 16 for the primary and secondary outcomes which corresponds to the maintenance phase. The center line shifted from 94% to 97% for the primary outcome, which compares with 93% versus 99% in the model based estimates discussed above. The center line shifted from 95% to 98% for the secondary outcome, which compares with 94% versus 98% in the model based estimates. The small multiples p charts are presented in the Supplemental Digital Content at <http://links.lww.com/PQ9/A136> for Figure. Variation is observed between and within groups, with some clinics still not at 100% performance by the conclusion of the intervention period and some clinics at 100% in the baseline period.

More than 90% of practices successfully met the process measure of >90% of providers with no unread or unacknowledged laboratory results in their EHR inbox for >72 hours by the fourth month of their intervention phase. Using the log-rank test, time to success on this process measure was not different between the 3 groups ( $P = 0.534$ ).

## DISCUSSION

In one of the first cluster randomized, stepped wedge trials to address pediatric ambulatory diagnostic-related errors,

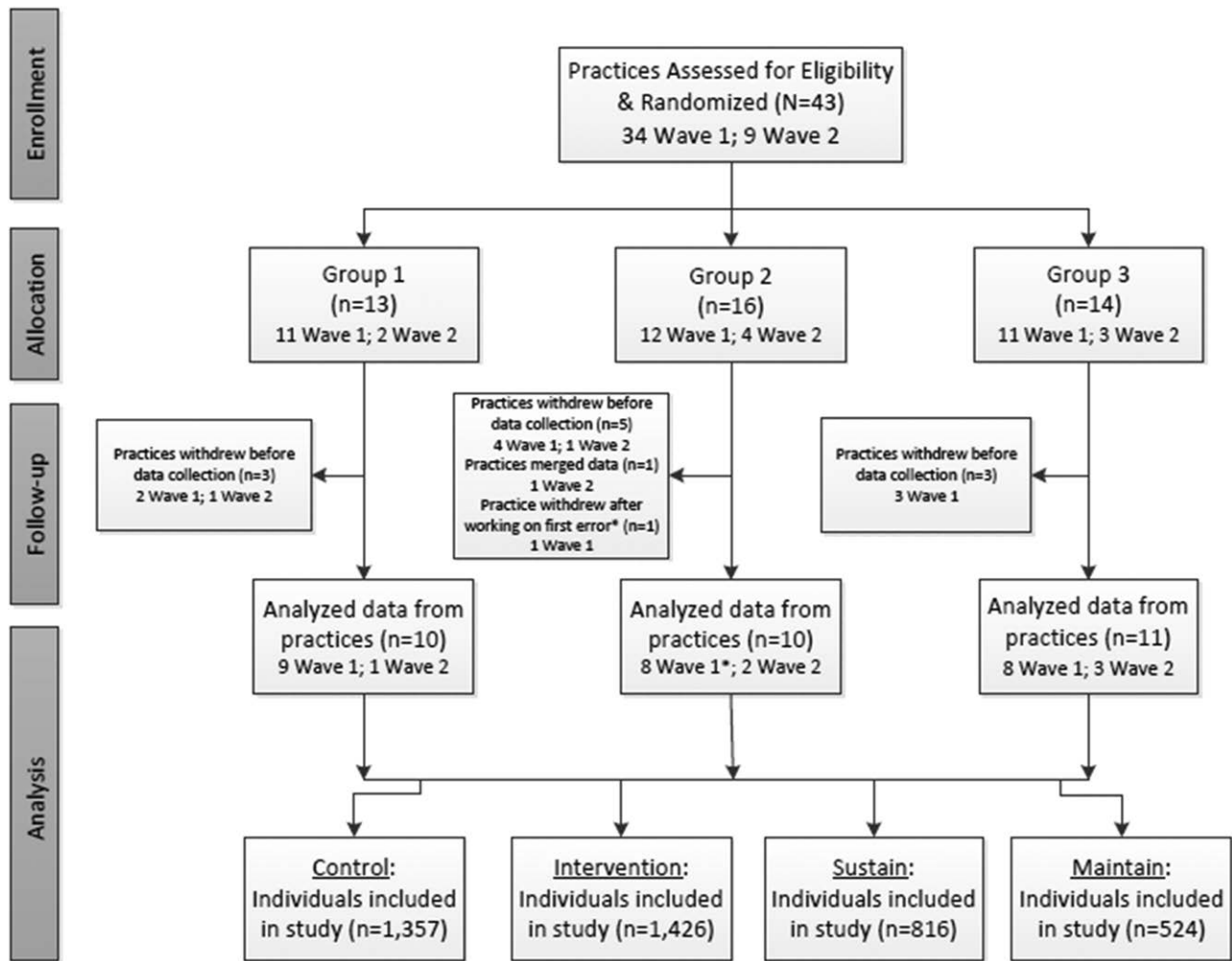


Fig. 2. Modified Consort Flow Diagram for Stepped Wedge Trial. \*1 practice withdrew after collecting control data for 8 months and working to reduce another error. We included their data for this phase.

Table 3. Demographics of Included Patients with Abnormal Laboratory Values in Primary Analysis

	Control	Intervention	Total	P
Demographics				
No. patients with abnormal laboratory values	1,357	1,426	2,783	
Age (1–2 years old)	53.7%	55.8%	54.7%	0.265
Female	51.1%	52.0%	51.6%	0.665
Private insurance*	23.7%	22.6%	23.1%	0.512
Abnormal Laboratory Values*				
Microcytic anemia	34.8%	41.1%	38.0%	<0.001
Elevated lead level	19.2%	15.4%	17.3%	0.008
Positive sexually transmitted infection	16.5%	18.9%	17.7%	0.098
Positive group A streptococcal throat culture	20.4%	15.0%	17.7%	<0.001
Elevated or depressed thyroid stimulating hormone	9.1%	9.6%	9.3%	0.688

\*One hundred twenty patients (4.3%) had missing insurance data and 75 patients (2.7%) had missing specific laboratory values and were excluded from primary analyses.

a national QIC intervention reduced the frequency of missed or delayed action on abnormal laboratory values in primary care pediatrics when comparing the sustain and maintenance versus control phases. The QIC failed to reduce error rates during the initial 8-month intervention phase for both the primary outcome appropriate action without a delay, or for the secondary outcome of action without a delay or documentation of laboratory

abnormality using both classical statistical methodologies and statistical process control chart. Significant reductions were appreciated in post hoc analyses comparing sustain and maintenance phases (months 9–24) to the control phase. These reductions are notable because practices were focusing QI efforts on other targets at those times. A potentially delayed effect might have resulted from process improvements that take time, achieving

**Table 4. Primary and Secondary Outcomes Results**

Outcome	Comparison	Total N	Model Based Percentages	RD (95% CI)	P
<b>Primary outcome</b>					
Percent of patients with a specific abnormal laboratory value* with a documented appropriate action without delay†	Intervention vs control	2,590	94.1% vs 93.0%	1.1% (−1.0%, 3.1%)	0.302
	Sustain vs intervention	2,085	95.9% vs 94.0%	1.9% (−0.5%, 4.2%)	0.118
	Maintenance vs sustain	1,251	95.1% vs 95.3%	−0.2% (−2.8%, 2.3%)	0.867
	Sustain vs control	2,029	96.5% vs 93.5%	3.0% (0.3%, 5.7%)	0.030
	Maintenance vs Control	1,756	98.8% vs 92.9%	5.9% (2.5%, 9.2%)	0.001
	Group 1 only: intervention vs control	815	91.0% vs 85.6%	5.4% (1.6%, 9.2%)	0.006
<b>Secondary outcome</b>					
Percent of patients with a specific abnormal laboratory value* where the provider documented that the result was abnormal, provided a diagnosis, and/or documented the appropriate action without delay†	Intervention vs control	2,590	94.5% vs 94.0%	0.1% (−1.8%, 1.9%)	0.922
	Sustain vs intervention	2,085	96.4% vs 95.1%	1.3% (−0.9%, 3.6%)	0.247
	Maintenance vs sustain	1,251	96.1% vs 95.9%	0.2% (−2.3%, 2.6%)	0.874
	Sustain vs control	2,029	96.4% vs 94.6%	1.8% (−0.7%, 4.3%)	0.149
	Maintenance vs control	1,756	98.3% vs 94.3%	4.0% (1.0%, 7.0%)	0.009
	Group 1 only: intervention vs control	815	91.6% vs 88.0%	3.6% (−0.1%, 7.3%)	0.054

Fitted by generalized mixed-effects model with potential clustering effects of practices and months taken into account; Adjusted for age, sex, insurance, and wave. Post-hoc analyses are presented below the dotted lines.

\*Abnormal laboratory values included (1) microcytic anemia, (2) elevated lead levels, (3) positive sexually transmitted infections, (4) positive group A streptococcal throat culture with negative rapid test, or (5) elevated or depressed thyroid stimulating hormone (TSH).

†Appropriate actions respectively included: (1) starting iron, sending iron studies, or family conversation on dietary iron; (2) family conversation on lead remediation or plan to retest; (3) antibiotics started or referral to HIV specialist; (4) antibiotics started or family conversation about positive test; (5) plan to repeat TSH test or referral to endocrinologist. Delays were no appropriate action within 30 days for microcytic anemia and elevated lead levels, and within 7 days for others.

>93% reliability might require more sophisticated-QIC approaches to improve test results management, and/or ascertainment bias may have led to higher control data for certain practices.

Despite evidence suggesting benefit of QICs,<sup>16,28–32</sup> this QIC’s effect may have appeared later because 8 months may not be sufficient to impact embedded workflows and long-standing processes and procedures. For example, managing test results in EHRs must consider prioritization of results (flagging abnormal or potentially dangerous results), electronic transmission of information, clear definition of responsibilities, training of providers to respond to alerts, and to document consistently to avoid communication failures.<sup>33</sup> Practices leveraged several EHR-based solutions and optimized protocols and policies. Some practices created a practice-wide EHR laboratory “inbox” to ensure teams did not miss abnormal values during provider absences. Practices also used EHR macros to make recall and documentation of actions easy for providers. These complex process interventions may require >8 months to reach maximal effectiveness.

Additionally, a QIC may need different types of intervention suggestions and change ‘toolkits’ to improve a process that is already more than 90% reliable, and/or sites may need to be stratified by baseline reliability level to identify relevant interventions. At level 2 reliability, appreciable constraints, affordances, differentiation of separate laboratory studies and “error-proofing” is likely required to see improvement.<sup>34</sup> Group 1, which had a lower model-based estimate control phase error rate than the aggregate of all 3 groups (86% vs 93%) did see a significant improvement in the intervention phase, which could support the idea that a ceiling effect contributed to a lack of improvement on the primary outcome. It is unclear if error rates in practices that chose to participate in a QIC are higher than the general population of pediatric

practices, possibly contributing to the ceiling effect. Prior research on QICs demonstrates improvement for processes with much lower reliability at baseline.<sup>16</sup>

Alternatively, given that Group 1 had a lower control phase error rate than the aggregate of all 3 groups, this could suggest ascertainment bias. Groups 2 and 3 were aware that they would ultimately work to reduce these errors and may have begun work before their intervention phase, thus accruing the intervention effect during the control phase. Anecdotal evidence for ascertainment bias includes practice teams reaching out for resources to reduce errors on which they were not yet assigned to work. The supplemental small multiples p charts further supports this conclusion as more clinics in Groups 2 and 3 reported 100% performance on the primary outcome in the baseline phase than clinics in Group 1. While the research team did not distribute project resources early, teams could work to reduce these errors on their own. Key interventions that practices mentioned included setting up procedures for checking laboratory values on holiday weekends and signing out laboratory value responsibilities when providers went on vacation. As opposed to our hypothesis above regarding EHR interventions requiring >8 months for full implementation, these interventions require less behavior change than the other errors addressed in Project RedDE and therefore could be implemented quickly and before formal training from the collaborative. This observation points to the fact that pediatric practices without a QIC infrastructure could implement these interventions and see positive results. The cluster-randomized, stepped wedge methodology allowed for improvement on 3 measures (labs, blood pressure, and depression) to be rigorously tested simultaneously. While this methodology increased the risk for ascertainment bias, it allowed for quicker improvement and understanding of whether QICs are beneficial for these errors.

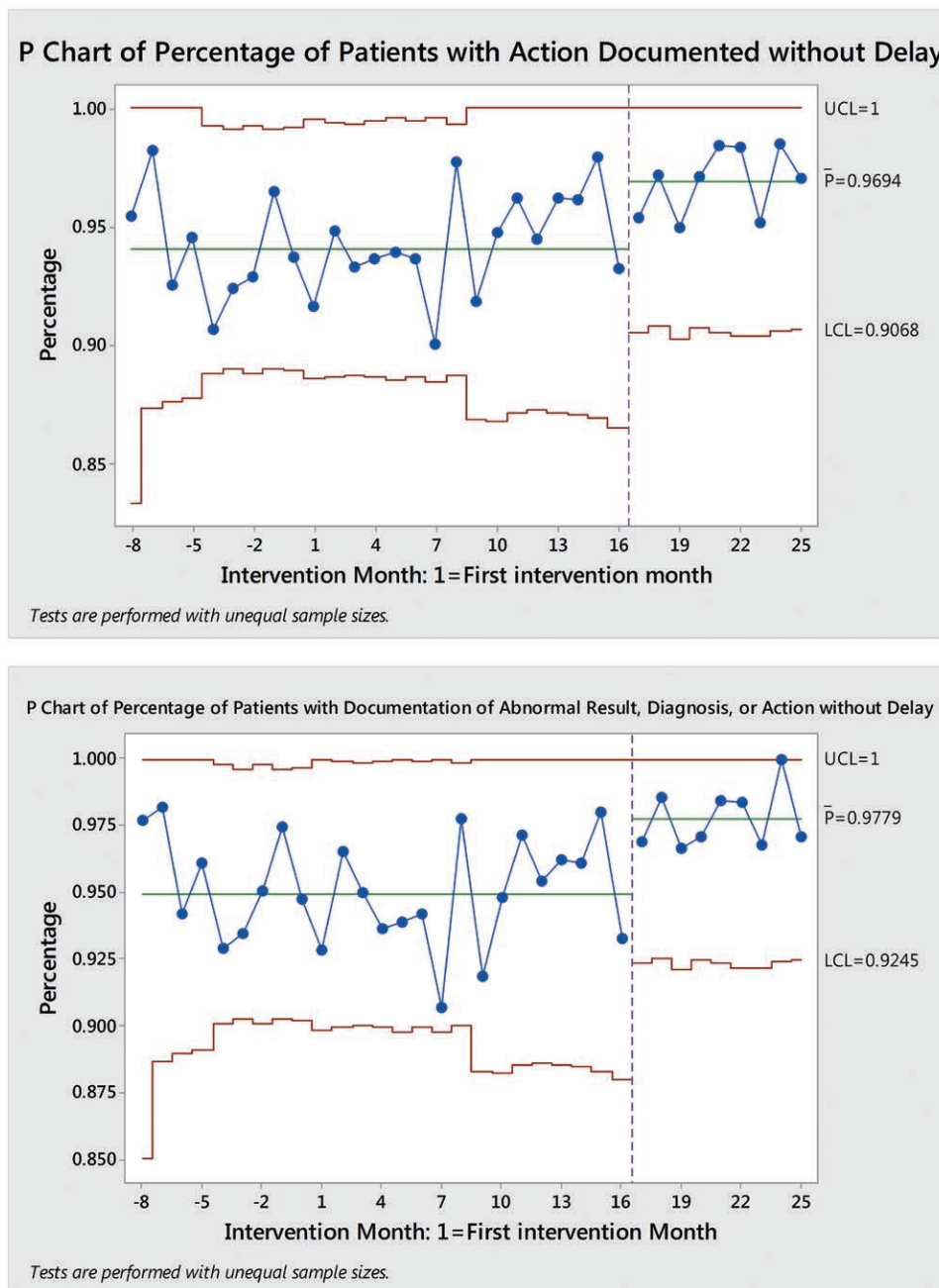


Fig. 3. P Charts of Primary and Secondary Outcomes.

For this reason, we believe that these rigorous methodologies should continue to be employed in QI research.

Interestingly, as opposed to our initial hypothesis, awareness or appropriateness of actions to take for these abnormal laboratory results did not seem to appreciably drive the primary outcome, as our secondary outcome which only required documentation of the laboratory abnormality was not noticeably higher (94.1% success in primary outcome vs 94.5% success in secondary outcome). Our work suggests the need for further implementation science research on the use of QIC interventions for complex, multi-step and multifaceted “sociotechnical” problems<sup>35</sup> such as follow-up of test results.

Our study has several limitations. Practices enrolled in a QIC to reduce errors are likely not representative of all pediatric practices. Our work likely did not impact other phases of the testing process, such as decision-making related to ordering tests in the pre-analytic phase.<sup>10</sup> Furthermore, appropriate actions on results were purposely broad, suggesting that some actions might be considered insufficient if examined more closely, although results did not appreciably change when documentation of abnormality was the outcome of interest. Error rates would be higher if we included all abnormal laboratory values. The small multiples p charts suggests that not all clinics improved equally, which presents further



opportunities for research into why some clinics were more or less responsive to the QIC intervention.

Additionally, the research team conducted no direct site visits or chart review verifications so there could be variability in the application of data definitions across practices despite the continuous availability of QuIN staff and the research team. Similarly, we are unable to speak to which aspects of the change package were implemented at each site and over what timeline. Doing so would have worsened the data collection burden, which already led to the attrition of 11 of the 43 practices randomized. The study does not have data on these 11 practices, preventing an intention to treat analysis or comparison of demographics. We believe challenges collecting EHR-based data at these 11 practices suggest a need for additional attention and resources to be devoted to primary care practice data collection from EHRs to facilitate future QI projects that are essential to improving our care delivery system. Finally, practices were asked to evaluate the first 10 patients' charts that met inclusion criteria from each month to reduce their burden for randomly selecting charts every month. While this is not a randomized assignment for chart review, it does create a quasi-randomized sampling strategy that is pragmatic and does not overburden clinic staff. We believe it is unlikely that this strategy is appreciably different from a randomized sampling strategy as patients who come on the first of the month, over the 29 month study duration, are likely not different from all patients. Nevertheless, the effects of such a quasi-random sampling strategy on the study outcomes, albeit likely little, are unknown.,

## CONCLUDING SUMMARY

Implementation of a QIC in a national group of United States pediatric practices reduced the frequency of missed or delayed action on abnormal laboratory values in analyses comparing sustain and maintenance phases to the control phase but not initial intervention to control phases. Future work should focus on understanding how a QIC functions in settings of 90% or more reliability at baseline and the time/effort required for improvement in already moderately highly reliable systems.

## ACKNOWLEDGEMENTS

The authors would like to gratefully acknowledge the contribution of all pediatric primary care sites who participated in this work and Drs. Brady, Kairys, Lilienfeld, O'Donnell, and Stein, as well as the American Academy of Pediatrics' Quality Improvement and Innovations Network, and Drs. Dadlez, Orringer, and Helms.

## DISCLOSURE

The authors have no conflict of interest to declare in relation to the content of this article.

## REFERENCES

1. Wahls TL, Cram PM. The frequency of missed test results and associated treatment delays in a highly computerized health system. *BMC Fam Pract.* 2007;8:32.
2. Singh H, Thomas EJ, Sittig DF, et al. Notification of abnormal lab test results in an electronic medical record: do any safety concerns remain? *Am J Med.* 2010;123:238–244.
3. Callen J, Georgiou A, Li J, et al. The safety implications of missed test results for hospitalised patients: a systematic review. *BMJ quality & safety.* 2011;20(2):194–199.
4. 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;145:488–496.
5. Callen JL, Westbrook JI, Georgiou A, et al. Failure to follow-up test results for ambulatory patients: a systematic review. *J Gen Intern Med.* 2012;27:1334–1348.
6. Eder M, Smith SG, Cappelman J, et al. *Improving Your Office Testing Process, A Toolkit for Rapid-Cycle Patient Safety and Quality Improvement.* Rockville, MD: Agency for Healthcare Research and Quality; 2013.
7. Poon EG, Gandhi TK, Sequist TD, et al. "I wish I had seen this test result earlier!": dissatisfaction with test result management systems in primary care. *Arch Intern Med.* 2004;164:2223–2228.
8. Callen J, Georgiou A, Li J, et al. The impact for patient outcomes of failure to follow up on test results. How can we do better? *EJIFCC.* 2015;26(1):38–46.
9. Singh H, Spitzmueller C, Petersen NJ, et al. Primary care practitioners' views on test result management in EHR-enabled health systems: a national survey. *J Am Med Inform Assoc.* 2013;20:727–735.
10. Hickner J, Graham DG, Elder NC, et al. Testing process errors and their harms and consequences reported from family medicine practices: a study of the American Academy of Family Physicians National Research Network. *Qual Saf Health Care.* 2008;17:194–200.
11. Ealovega MW, Tabaei BP, Brandle M, et al. Opportunistic screening for diabetes in routine clinical practice. *Diabetes Care.* 2004;27:9–12.
12. Rinke ML, Singh H, Ruberman S, et al. Primary care pediatricians' interest in diagnostic error reduction. *Diagnosis.* 2016;3(2):65–69.
13. Rinke ML, Singh H, Heo M, et al. Diagnostic errors in primary care pediatrics: project RedDE. *Acad Pediatr.* 2018;18(2):220–227.
14. Bundy DG, Singh H, Stein RE, et al. The design and conduct of project RedDE: a cluster-randomized trial to reduce diagnostic errors in pediatric primary care. *Clin Trials.* 2019;1740774518820522.
15. Greevy R, Lu B, Silber JH, et al. Optimal multivariate matching before randomization. *Biostatistics.* 2004;5(2):263–275.
16. Schouten LM, Hulscher ME, van Everdingen JJ, et al. Evidence for the impact of quality improvement collaboratives: systematic review. *BMJ.* 2008;336:1491–1494.
17. Rinke ML, Chen AR, Bundy DG, et al. Implementation of a central line maintenance care bundle in hospitalized pediatric oncology patients. *Pediatrics.* 2012;130:e996–e1004.
18. Rinke ML, Bundy DG, Chen AR, et al. Central line maintenance bundles and CLABSIs in ambulatory oncology patients. *Pediatrics.* 2013;132(5):e1403–1412.
19. (ONC) TOotNCfHIT. Safety Assurance Factors for Electronic Health Record Resilience (SAFER) Guides. 2014. Available at <https://www.healthit.gov/topic/safety/safer-guides>. Accessed May 24, 2018.
20. Rinke ML. Toolkit for Reducing Diagnostic Errors in Primary Care - Project RedDE! 2018. Available at <https://www.aap.org/en-us/professional-resources/quality-improvement/Project-RedDE/Pages/Project-RedDE.aspx>. Accessed November 15, 2018.
21. American Academy of Pediatrics Committee on Environmental H. Lead exposure in children: prevention, detection, and management. *Pediatrics.* 2005;116(4):1036–1046.
22. Dahlberg RL. *Preventing childhood lead poisoning in New Jersey: Advocates and state government working together to increase the lead screening of children.* ACLU Foundation. 2005.
23. Nusbaum MR, Wallace RR, Slatt LM, et al. Sexually transmitted infections and increased risk of co-infection with human immunodeficiency virus. *J Am Osteopath Assoc.* 2004;104:527–535.
24. Ku L, St Louis M, Farshy C, et al. Risk behaviors, medical care, and chlamydial infection among young men in the United States. *Am J Public Health.* 2002;92:1140–1143.
25. Ginocchio RH, Veenstra DL, Connell FA, et al. The clinical and economic consequences of screening young men for genital chlamydial infection. *Sex Transm Dis.* 2003;30:99–106.

26. Barash J. Group A streptococcal throat infection - to treat or not to treat? *Acta Paediatr.* 2009;98:434–436.
27. Nelson LS. The shewhart control chart-tests for special causes. *J Qual Technol.* 1984;16(4):237–239.
28. Miller MR, Niedner MF, Huskins WC, et al.; National Association of Children's Hospitals and Related Institutions Pediatric Intensive Care Unit Central Line-Associated Bloodstream Infection Quality Transformation Teams. Reducing PICU central line-associated bloodstream infections: 3-year results. *Pediatrics.* 2011;128:e1077–e1083.
29. Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med.* 2006;355:2725–2732.
30. Nadeem E, Olin SS, Hill LC, et al. Understanding the components of quality improvement collaboratives: a systematic literature review. *Milbank Q.* 2013;91:354–394.
31. Hulscher ME, Schouten LM, Grol RP, et al. Determinants of success of quality improvement collaboratives: what does the literature show? *BMJ Qual Saf.* 2013;22:19–31.
32. Beers LS, Godoy L, John T, et al. Mental health screening quality improvement learning collaborative in pediatric primary care. *Pediatrics.* 2017;140(6):pii: e20162966.
33. Sittig DF, Singh H. Improving test result follow-up through electronic health records requires more than just an alert. *J Gen Intern Med.* 2012;27:1235–1237.
34. Nolan T, Resar R, Haraden C, et al. *Improving the Reliability of Health Care.* Cambridge, MA: Institute for Healthcare Improvement; 2004.
35. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. *Qual Saf Health Care.* 2010;19(Suppl 3):i68–i74.