1. TITLE PAGE

Reducing Diagnostic Errors in Primary Care Pediatrics

Principal Investigator: Michael L. Rinke, MD, PhD

Co-Investigators: David G. Bundy, MD, Hardeep Singh, MD, MPH, MPH, Moonseong Heo, PhD, Jason S. Adelman, MD, MS, Heather C. O'Donnell, MD, MS, Vanessa Shorte, Amanda Norton

Expert Group: Ruth E.K. Stein, MD, Tammy M. Brady, MD, MHS, Christoph U. Lehmann, MD, Nina Dadlez, MD, MSc, Kelly Orringer, MD

Organization: Albert Einstein College of Medicine

Dates of Project: 9/30/2014 – 9/29/2019

Federal Project Officer: Monika Haugstetter

Acknowledgement of Agency Support: The investigators gratefully acknowledge the support of the Agency for Healthcare Research and Quality.

Grant Number: 5R01HS023608

2. STRUCTURED ABSTRACT

Purpose: To determine whether a quality improvement collaborative (QIC) is associated with a reduction in three specific diagnostic error rates in a national group of pediatric primary care practices.

Scope: Primary care pediatric practices

Methods: During a cluster-randomized clinical trial, practices worked in different orders to improve performance on each of three different diagnoses: elevated blood pressure (BP), adolescent depression, and abnormal laboratory values. While improving their first diagnosis during an 8-month action period, practices collected control data for a different diagnosis. In two subsequent 8-month periods, practices worked to improve two additional diagnoses and continued to provide data on the ability to sustain and maintain improvements. The QIC intervention included day-long video conferences, transparent data sharing, analysis of failures, QI coaching, and tools to help improve diagnostic performance. The primary outcomes were the measured frequency of actions taken or depression diagnoses in control versus intervention conditions, compared via generalized mixed-effects regression models.

Results: Forty-three practices were randomized with 31 in the final analysis. Comparing control versus intervention phases, the mean adjusted percentage of patients who received appropriate BP actions increased from 58% to 74% (risk difference (RD) 16%; 95% CI 12%, 20%). Patients with depression diagnoses increased from 6.6% to 10.5% (RD 3.9%; 95% CI 2.4%, 5.3%). Patients who received appropriate laboratory actions did not change (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%).

Key Words: Ambulatory, diagnostic errors, pediatric

3. PURPOSE

The specific aim and purpose of this grant was as follows:

Specific Aim: To determine whether a quality improvement collaborative consisting of evidence-based best-practice methodologies, mini root cause analyses, data sharing, and behavior change techniques is associated with a reduction in three specific diagnostic error rates in a national group of pediatric primary care practices.

> *Hypothesis 1: Implementation of a quality improvement collaborative will lead to a 40% reduction in missed diagnosis of adolescent depression.*
> *Hypothesis 2: Implementation of a quality improvement collaborative will lead to a 30% reduction in missed diagnosis of pediatric elevated blood pressure.*
> *Hypothesis 3: Implementation of a quality improvement collaborative will lead to a 45% reduction in delayed diagnosis of actionable laboratory results.*

4. SCOPE

The Institute of Medicine (IOM) report "*Improving Diagnosis in Health Care*" highlights the significance of diagnostic errors (DE) and defines them as "the failure to establish an accurate and timely explanation of the patient's health problem(s) or communicate that explanation to the patient."[1] The report asserts that each of us will likely have a meaningful DE in our lifetime, with one estimate suggesting that DEs affect 1 in 20 outpatient adults annually.[2] DEs are also responsible for approximately $34 billion dollars in annual United States malpractice payments.[3] Although studies on reducing ambulatory diagnostic breakdowns in adults have emerged,[4,5] little progress has been made to understand or reduce ambulatory pediatric DEs.[6]

In surveys, 35-54% of pediatricians reported a DE occurring at least monthly and 33-45% reported DEs that harmed a patient at least annually.[7,8] The true burden is likely higher, given that physicians generally underestimate their personal error rates.[9] Research into pediatric ambulatory DEs remains in its infancy, and it is additionally compounded by challenges in defining and measuring DEs.[10] This study seeks to identify these errors across a broad range of pediatric ambulatory clinics and then reduce them through quality improvement collaborative methodology.

When choosing which DEs to address, primary care pediatricians expressed more interest in working to reduce high-frequency/sub-acute DEs, such as missed hypertension diagnosis, versus low-frequency/acute DEs, such as missed appendicitis diagnosis.[8] The epidemiology of these sub-acute DEs is unknown, but their high frequency and long-term health effects may lead to increased morbidity and cost compared to low-frequency/acute DEs. For example, adolescent depression affects 20% of adolescents before age 20, and 7.8% of adolescents attempt suicide.[11-16] Unfortunately, in 60 children with probable mental health diagnoses, only 15 (25%) were identified by pediatricians, and in only 14% did pediatricians consider a psychiatric referral.[17] Similarly, 3-5% of children have hypertension.[18,19] The first step in diagnosing hypertension is recognizing when blood pressure (BP) is elevated. Recognition in both the hypertensive and pre-hypertensive range is important, as pre-hypertensive children are at greater risk for developing hypertension and have worse cardiovascular outcomes when compared to normotensive children.[20,21] In a single-center study, 39% of pediatric visits included an elevated BP, but only 13% of these elevations were recognized by providers.[22] Finally, 40% of ambulatory primary care visits include laboratory testing,[23] but 83% of physicians report at least one delay in reviewing laboratory results during the previous 2 months,[24] and 40% report missing results despite a highly computerized health system.[25] It is crucial to investigate the epidemiology of these high-frequency/sub-acute DEs across multiple practices to better describe the pervasiveness of pediatric DEs, as potentially many more patients are affected by this type of error. Rigorous epidemiologic, multi-site studies can also increase generalizability of findings, demonstrate models for other DE measures, and create an imperative to reduce these errors.

Our objective, via a prospective, stepped-wedge, cluster-randomized controlled trial in a national cohort of pediatric primary care clinics, was to determine whether a quality improvement collaborative (QIC) intervention could reduce the frequency of missed elevated pediatric BP, missed adolescent depression, and delayed action on laboratory values and sustain reductions while practices refocused on reducing other errors. Lessons from this project can be applied broadly and serve as a foundation for hypertension care delivery improvement efforts.

5. METHODS

Project RedDE (Reducing Diagnostic Errors in Pediatric Primary Care) aimed to reduce three different DEs in primary care pediatric practices in collaboration with the American Academy of Pediatrics' (AAP) Quality Improvement Innovation Networks (QuIIN) via a QIC. QuIIN aims to "improve the quality and value of care and outcomes for children and families" via quality improvement networks. QICs are an organized, multifaceted approach to QI with 1) a specific topic 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 targets for improvement, data feedback to teams, and small tests of change; and 5) a series of structured activities to advance improvement, exchange ideas and share experiences of participating teams.[26-31] Reducing missed elevated pediatric BP, reducing missed adolescent depression, and reducing delayed action on abnormal laboratory values were the three errors addressed by Project RedDE's QIC.

Recruitment and Randomization

In March 2015, we recruited 34 pediatric practices via email listserves and orientation webinars and randomized them via computer random number generator, in a non-blinded fashion, to one of three groups. We employed multivariate matching before randomization[32] based on university affiliation, the presence of a self-reported prior record of working to reduce the target DEs, and total annual visits per total number of pediatricians or nurse practitioners in the clinic. Nine practices dropped out after randomization but before submitting data due to inability to collect data. Of the remaining 25 practices, 24 submitted complete project data through September 2017; one practice dropped out after 8 months when their lead physician left the practice. We included this practice's data in analyses of depression and laboratory errors, as they submitted data for those errors but not for BP. Nine additional practices were recruited in December 2015 to increase the size of the cohort and were similarly randomized via computer random number generator in a non-blinded fashion. Of these, two practices dropped out after randomization but before submitting data, also due to data burden; two other practices from a single care network merged into one team to boost their practice sample size. These six 'Wave 2' teams participated alongside the 24 'Wave 1' teams. In this manner, we randomized 43 total practices and included 30 in the final analysis.

Study Design

In July 2015, each of the three groups was assigned to collect retrospective baseline data (February-June 2015) on one of three DEs: missed elevated BP, delayed diagnosis of abnormal laboratory values, or missed diagnosis of adolescent depression.[33] *A priori*, each error was to be examined independently. The groups collected 1 month of prospective baseline data (September 2015) and then began an 8-month QI action period in October of 2015 to reduce their assigned error. Concurrently (September 2015-May 2016), each group collected control data on a second DE. In a prospective, stepped-wedge fashion, after 8 months, (June 2016) each group began to work to reduce a second DE during a second action period, sustain the improvement on their first error, and collect control data for the third DE. In February 2017, each group began to work to reduce the third error during a third action period, sustain the improvement on their second error, and maintain the improvement on their first error, with reduced feedback and attention on the first error from the larger QIC (Figure 1).

Using this design, each group of practices had a 'control phase,' when they collected data on all errors but did not attempt to reduce them, and all but one Wave 2 group had an 'intervention phase,' when they actively worked to reduce all errors. Groups also had a 'sustain phase,' when they actively worked to reduce a second DE and sustain improvement on their first error; finally all groups had a 'maintenance phase,' when they actively worked to reduce two other DEs and maintain improvement on their first error.

Intervention

The primary intervention was a QIC. Each practice identified a three-person QI team consisting of a physician, a nurse, and another professional (e.g., administrator, business associate, front desk staff, etc.). After completing baseline data collection, teams participated in a 2-day video conference, where they learned and practiced QI methodology and DE-specific content. Though all teams participated in the QIC video conference, only the teams about to intervene on a given error received information and training on this error. Following this, teams received rapid, transparent data feedback on performance with benchmarking, participated in monthly hour-long video conferences, and completed monthly mini root cause analyses. These mini root cause analyses examined a patient with a BP error in their clinic and 15 standardized patient and systems factors that could have led to this error.[34,35] Teams focused their video conferences and mini root causes analyses on their specific error while in that error's intervention phase. 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 conference learning sessions were conducted every 8 months as practices transitioned to working on a new DE (Figure 1). When practices were working on their second DE, monthly video conferences provided transparent data feedback in the form of run-charts from both their first and second DEs. When working on their third DE, monthly video conferences presented data from both their second and third DEs, and data from their first error were only presented quarterly. Practices could always access all of their data independently. We believe practices spent an average of 4 hours per month working on Project RedDE-related activities: 8-hour learning session every 8 months, 1-hour video conference every non-learning session month, 1 hour for team quality improvement meetings, and 1 hour for data entry and collection. In addition, practices spent time developing and implementing changes, including new tools and workflows, for which t theime spent cannot be easily estimated. Figure 1 presents the process flow for practices working to reduce BP errors. Similar process flows existed for depression and laboratory values.

Figure 1: Project RedDE Timeline for Missed Elevated Blood Pressure (BP)

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
Wave 1							
Group 1	Control Phase	Virtual Learning Session 1	Intervention Phase	Virtual Learning Session 2	Sustain Phase [b]	Virtual Learning Session 3	Maintenance Phase
Group 2	[a]		Control Phase		Intervention Phase		Sustain Phase
Group 3	[a]		[a]		Control Phase		Intervention Phase
Wave 2[c]			**November 2015-March 2016**				
Group 1			[a]		[a]		Control Phase
Group 2			Control Phase		Intervention Phase		Sustain Phase
Group 3			[a]		Control Phase		Intervention Phase

Similar process flows existed for Depression and Laboratory Errors.

[a] Practices were involved in Project RedDE during this time but working exclusively on the two non-BP errors. Practices in Groups 2 and 3 had already worked to reduce 1 or 2 other diagnostic errors respectively before beginning to work on BP errors.

[b] During the Sustain and Maintenance Phases, practices began working to reduce a second and third diagnostic error respectively.

[c] 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.

Project leadership developed change packages for each error with the aim of helping teams 1) implement uniform processes; 2) use systematic tools to identify patients at risk for these diagnoses; and 3) help providers know, perform, and document appropriate actions.

All these resources were maintained on the Project RedDE website, and practices shared new and modified tools throughout the QIC. All resources were made available to the public following the project's conclusion as one of the grant's deliverables in a AAP Toolbox.[36]

Measures

We utilized pragmatic error measures with efficient data collection methods to accommodate the needs of high-throughput practices.

Inclusion criteria for the elevated BP DE measure were patients ≥ 3 years old through 22 years old who had an elevated systolic or diastolic BP recorded at their health supervision visit. We defined elevated BPs as ≥ 90th percentile for age, height, and sex or ≥ 120 mmHg systolic or 80 mmHg diastolic pressure at any age.[37] The primary outcome measure was the number of patients with an elevated BP with an appropriate action taken by the provider per 100 patients with elevated BP. This provider "appropriate action" confirms that a diagnosis was made, as not all providers document a diagnosis of 'an elevated BP.' Appropriate actions included any of the following: a) rechecking the BP, b) noting a plan to recheck the BP at a future visit, c) referral to a hypertension specialist (e.g., pediatric cardiologist or nephrologist), and/or d) laboratory or radiologic studies ordered to evaluate causes of elevated BP. More than one action could be selected, and actions had to occur within 30 days of the visit.

Definitions of “appropriate actions” were necessarily broad as the study relied on frontline clinicians with limited time to collect data. A research team chart review was beyond the scope of this work.

Although depressive symptoms (e.g., poor school performance, interrupted sleep patterns, increased disruptive behaviors, etc.) without appropriate provider identification or referral were considered as the primary outcome, pilot data suggested that the number of times a patient has documented signs and symptoms of depression but is not referred to or already receiving mental health treatment is rare, and this methodology likely underidentifies adolescent depression.[38] Thus, we used a proxy primary outcome measure for poor adolescent depression diagnostic performance: the frequency of adolescent depression recognition and diagnosis, which increases as missed diagnoses decrease. Given prior literature,[17,38,39] it is reasonable to assume an underdiagnosis of adolescent depression. Practices identified the percent of adolescents who carried diagnoses of depression, dysthymia or sub-syndromal depression in visit notes, problem lists or billing records (International Classification of Diseases(ICD)-9/10 codes 296.2, 296.3, 311, 311.0, 300.4, 309.0, 309.1, 309.4; F32.0-5, F32.9, F33.0-4, F33.8-9, F34.1, F43.21, F43.25, F06.3X). These were not patients who only screened positive on a PHQ-9M screen but those who were ultimately given, at that visit or within 30 days of the visit, a diagnosis of depression. Sub-syndromal depression was defined as “a depressive state having two or more symptoms of depression of the same quality as in major depression(MD), excluding depressed mood and anhedonia.”[40] We included patients if they were 11 to 23 years old and attending a health supervision visit. We included 11-year old patients because the AAP recommended screening at this age, although the current recommendation is to start screening for 12-year-old patients.[41] Charts were checked 30 days after the visit.

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*
Microcytic anemia	Hemoglobin <11g/dL and Mean corpuscular volume <75fL	1 and 2 year olds	Starting iron supplementation, sending iron studies, or family conversation on increasing dietary iron intake	30 days
Elevated lead level	Blood lead >5µg/dL	1, 2, and 3 year olds	Family conversation on lead remediation or plan to retest for elevated lead levels	30 days
Sexually transmitted disease	Positive *Neisseria gonorrhoeae, Chlamydia trachomatis, Treponema pallidum,* or human immunodeficiency virus	> 10 years old	Antibiotics started or referral to infectious disease specialist	7 days
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 days
Possible hypo- or hyperthyroid	Thyroid stimulating hormone (TSH) <0.5µIU/mL or >4.5µIU/mL	≥ 1 year old	Plan to repeat TSH test or referral to endocrinologist	7 days

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

For laboratory values, the primary outcome was the proportion of patients with any of five specific abnormal laboratory values who had appropriate action documented without delay per 100 patients. These five 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.[42-47] Definitions of “appropriate actions” and delays, 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 (Table 1).

Practices were taught measure definitions via multiple webinars, slides, and written materials. Listservs and QI coaches were available for questions and clarifications. For each eligible patient, practices recorded age, sex, and insurance status (public, private, self, unknown) and entered data into a web-based portal. Insurance status was included as a potential confounder, because it is an easily collectible, partial marker of socioeconomic status, which has previously been shown to be associated with errors in ambulatory care.[48]

Practice demographics, including items such as university affiliation, previous work on these errors, clinic and patient demographics, and QI skill, were identified via self-report questionnaire before the start of the project.

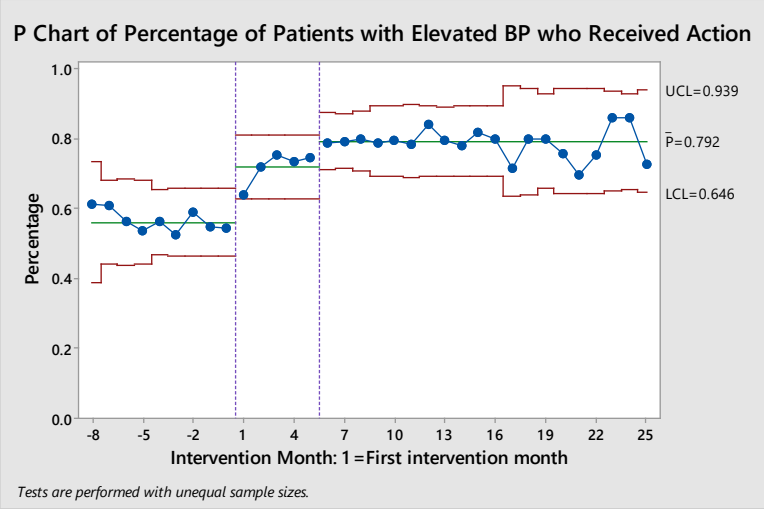
Statistical Analysis:

We used patients as the unit of analysis and compared the primary outcome of a mean number of patients with an appropriate provider action taken per 100 patients who met inclusion criteria between the intervention and control phases. The primary outcome effect measures are presented as model-based estimates of risk differences (RD). We applied generalized mixed-effects logistic regression models adjusted for age, sex, insurance status, and wave with month-specific and practice-specific intercepts considered random, whereas age, sex, and insurance status were considered fixed. We excluded patients with incomplete demographic data from the final analysis. We completed all data analyses with SAS v9.3. This study was approved by the AAP's and the Albert Einstein College of Medicine's Institutional Review Boards.

6. RESULTS

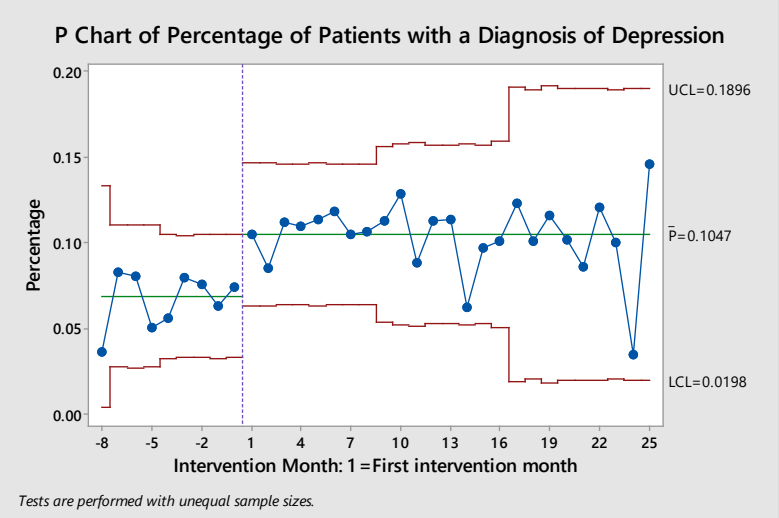
Principal Findings

BP:



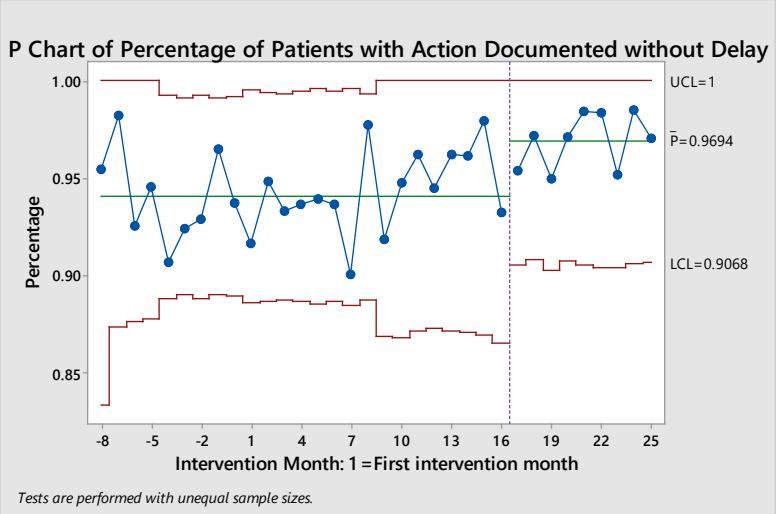
Data on 1,728 patients were available for the control phase and on 1,834 patients for the intervention phase. We excluded 140 patients (3.9% of total patients) from the final analysis due to missing insurance data. The model-based estimated mean percentage of patients with either elevated systolic or diastolic BP who received an appropriate action increased from 57.6% in the control phase to 73.5% in the intervention phase (RD 16.0%; 95% CI 12.3%, 20.0%, p<0.0001). Of the 1,366 intervention and 969 control patients who received an appropriate action, 84% had their BP rechecked in the intervention phase versus 75% in the control phase (p<0.001); 27% had a plan to recheck BP at a future visit in the intervention phase versus 22% in the control phase (p=0.004); and 3% had a referral to a specialist in the intervention phase versus 7% in the control phase (p<0.001). Practices continued to improve comparing the intervention and sustain phases (RD 5.2%; 95% CI 1.5%, 8.9%; p=0.006) and neither worsened nor improved comparing the maintenance and sustain phases (RD 0.9%; 95% CI -4.7%, 6.6%; p=0.743).

Depression:



Data on 3,394 patient visits were entered for the control phase and 4,114 for the intervention phase. We excluded 295 patients from the final model due to missing insurance data. There were statistically younger and more non-privately insured patients in the intervention phase. The adjusted percentage of patients with depression, dysthymia or sub-syndromal depression in the control phase was 6.6%, compared to 10.5% in intervention phase (RD 3.9%; 95% CI 2.4%, 5.3%; p<0.0001). Practices sustained and maintained these improvements: the mean percentage of patients with depression was not different when comparing the intervention to the sustain phase (RD -0.4%; 95% CI -2.3, 1.4%; p=0.642) or the sustain to the maintenance phase (RD -0.1%; 95%CI -2.7%, 2.4%; p=0.911).

Abnormal Laboratory Values:



Data on 1,357 patients were available for control and on 1,426 patients for the intervention phase. Due to missing insurance data or missing laboratory test data, we excluded 193 patients (7%). 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).

In *post hoc* analyses comparing sustain and control phases as well as maintenance and control phases, practices significantly improved (respectively, 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). 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).

Discussion:

In one of the first cluster-randomized, stepped-wedge trials to address pediatric or adult DEs, a national QIC intervention increased recognition of elevated BP in primary care pediatrics by 28% from baseline---an increase that was sustained for 16 months even when practices began focusing QI efforts elsewhere. Similarly, the QIC successfully increased the percent of adolescents who carried diagnoses of depression from 6.6%, to 10.5% and sustained this improvement over 16 months.

The QIC failed to reduce missed or delayed action on abnormal laboratory values error rates during the initial 8-month intervention phase for the primary outcome appropriate action without delay, 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 > 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.

Although missing elevated BP is not immediately dangerous, this error is high frequency[33]; primary care pediatricians would like to see it prioritized,[8] and it eventually takes a significant toll on pediatric and adult health.[49,50] Reducing common and potentially harmful errors through collaboration, data benchmarking, QI coaching and mini root cause analyses offers one possible path for DEs, for which few interventions focus on pediatric and ambulatory patients.[6] It is unclear if improvement came from the bundle of intervention tools provided to the practices or the focus on elevated BP that came from being part of a national QIC. By focusing on just one error at a time, practices would also likely experience less data collection burden, reducing the risk of attrition. The effect size of results seen in Project RedDE are comparable to previous QIC results,[30] especially when considering studies with a comparable initial prevalence of errors.[51] QICs are often resource intensive and Project RedDE demonstrates a benchmark for what practices across the country can achieve with dedicated focus and collaboration. The low attrition rate once practices were able to demonstrate data collection capacity (one of 31 practices) suggests that the burden of participating and working to improve these errors was not overwhelming and that practices found value in this work.

Missed opportunities to diagnose depression occur in approximately 60% of adolescents,[52] and such high-frequency errors are a priority for pediatricians.[8] Our data support prior studies suggesting under-diagnosis of adolescent depression is common in pediatrics,[17,38,39] and illustrate a methodology to reduce these misdiagnoses through collaboration, data benchmarking, QI coaching, and focusing on failures. Systematizing office practices to ensure screening with the PHQ-9M may improve diagnosis rates, as this process measure increased with depression diagnoses. Although measuring time to treatment and symptom relief was beyond the scope of this project, many practices anecdotally reported increased confidence in managing mild to moderately depressed adolescents, increased communication and collaboration with mental health practitioners, and improved outcomes for patients they otherwise would not have suspected of having depression. Given the sustainability of improvements when practices were focused on other diagnoses, we hypothesize that the change seen can be replicated in practices without an extensive QIC infrastructure, because improvements were consistent when the QIC was not focused on depression diagnosis. Further work is needed to understand why some clinics improved immediately and some clinics did not see appreciable improvement across the intervention phase.

Despite evidence suggesting benefit of QICs,[26-31] this QIC's effect may have appeared later, because 8 months may not be sufficient to impact embedded workflows and longstanding 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, and training of providers to respond to alerts and to document consistently to avoid communication failures.[53] Practices leveraged several EHR-based solutions and optimized protocols and policies. Some practices created a practice wide EHR laboratory "inbox" to ensure that 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 more than 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" are likely required to see improvement.[54] Group 1, which had a lower model-based estimate control phase error rate than the aggregate of all three groups (86% versus 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.

Limitations:

Limitations of this study include the concern that practices enrolled in a QIC are unlikely to be representative of all pediatric practices. Furthermore, 11 of the 43 randomized practices withdrew before study implementation due to data collection burden. All these practices withdrew before attempting to change their clinic processes and behaviors, making it unclear if easier data collection would have reduced this attrition rate. The study does not have information on practices that dropped out either for an intention-to-treat analysis or to compare demographics, as these practices did not submit any data. Practices with more resources or abilities to collect needed data may be less likely to resemble other general pediatric practices. However, we believe this to be not likely, as our cohort included great diversity, with single-practitioner private practices and large academic practices with many residents and attending physicians. Additionally, because the research team performed no direct site visits, there was potential variability in the application of data definitions across practices. However, the research team was available to answer questions during all data collection phases, hosted review sessions, and shared tips frequently on the listserv. Furthermore, appropriate actions on elevated BP and abnormal laboratory values 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. Small multiples p charts suggest 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.

Implications/Conclusions:

Implementation of a QIC in a national group of United States pediatric practices reduced the frequency of elevated BP and adolescent depression DEs in primary analyses. Missed or delayed action on abnormal laboratory values DEs were not reduced in primary analyses but were reduced in analyses comparing sustain and maintenance phases to the control phase. Future research should focus on spreading this effort to all pediatric primary care clinics, on the outcomes of patients following these diagnoses, and whether this model can apply to other diagnoses in primary care. Additionally, implementation science work could 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.

7. LIST OF PUBLICATIONS and PRODUCTS

1. Heo M, Kim N, Rinke ML, Wylie-Rosett J. *Sample size determinations for stepped-wedge clinical trials from a three-level data hierarchy perspective*. Statistical Methods in Medical Research. 2018 27:480-489.

2. Rinke ML, Singh H, Heo M, Adelman J, O'Donnell HC, Choi SJ, Norton A, Stein REK, Brady TM, Lehmann CU, Kairys SW, Rice-Conboy E, Thiessen K, Bundy DG. *Diagnostic errors in primary care pediatrics: Project RedDE*. Academic Pediatrics. 2018. 18:220-227.

3. Bundy DG, Singh H, Stein REK, Brady TM, Lehmann CU, Heo M, O'Donnell HC, Rice-Conboy E, Rinke ML. *The design and conduct of Project RedDE: A cluster-randomized trial to reduce diagnostic errors in pediatric primary care*. Clinical Trials. 2019. Feb 5. [Epub ahead of print]

4. Douglas LC, Heo M, Azad N, Racine A, Rinke ML. *Contextual factors associated with quality improvement success in a multi-site ambulatory setting.* Journal for Healthcare Quality. 2019. 41:317-328.

5. Rinke ML, Singh H, Brady TM, Heo M, Kairys SW, Orringer K, Dadlez NM, Bundy DG. *Cluster randomized trial reducing missed elevated blood pressure in pediatric primary care: Project RedDE.* Pediatric Quality and Safety. 2019. Sept 30. [Epub ahead of print]

6. Rinke ML, Bundy DG, Stein REK, O'Donnell HC, Heo M, Sangvai S, Lilienfeld H, Singh H. *Increasing recognition and diagnosis of adolescent depression: Project RedDE a cluster randomized trial.* Pediatric Quality and Safety. 2019. Sept 30. [Epub ahead of print]

7. Rinke ML, Bundy DG, Lehmann CU, Heo M, Adelman JS, Norton A, Singh H. *Project RedDE: Cluster randomized trial to reduce missed or delayed abnormal laboratory value actions.* Pediatric Quality and Safety. 2019. Sept 30. [Epub ahead of print]

8. Dadlez NM, Adelman JS, Bundy DG, Singh H, Rinke ML. *Contributing factors to pediatric ambulatory diagnostic errors: Project RedDE*. Pediatric Quality and Safety. 2020. [Submitted with Revisions Requested by Journal]

9. Douglas L, Bundy DG, Heo M, Singh H, Rice-Conboy E, Rinke ML*. Associations between contextual factors and project success in a multi-center quality improvement collaborative.* 2020. [Submitted for publication].

10. Rinke ML and American Academy of Pediatrics. *Toolkit for Reducing Diagnostic Errors in Primary Care - Project RedDE!* Available at: https://www.aap.org/en-us/professional-resources/quality-improvement/Project-RedDE/Pages/Project-RedDE.aspx

References:

1. National Academies of Sciences Engineering and Medicine. *Improving Diagnosis in Health Care.* Washington, DC: The National Academies Press; 2015.
2. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ quality & safety.* 2014;23(9):727-731.
3. Tehrani AS, Lee H, Mathews S, Shore A, Frick KD, Makary M. 20-Year summary of U.S. malpractice claims for diagnostic errors from 1985-2005 [Abstract]. 33rd Annual Meeting of the Society for Medical Decision Making; October 22-26, 2011; Chicago, IL.
4. Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care--a systematic review. *Family practice.* 2008;25(6):400-413.
5. Singh H, Weingart SN. Diagnostic errors in ambulatory care: dimensions and preventive strategies. *Advances in health sciences education : theory and practice.* 2009;14 Suppl 1:57-61.
6. McDonald KM, Matesic B, Contopoulos-Ioannidis DG, et al. Patient safety strategies targeted at diagnostic errors: a systematic review. *Ann Intern Med.* 2013;158(5 Pt 2):381-389.
7. Singh H, Thomas EJ, Wilson L, et al. Errors of diagnosis in pediatric practice: a multisite survey. *Pediatrics.* 2010;126(1):70-79.
8. Rinke ML, Singh H, Ruberman S, et al. Primary Care Pediatricians' Interest in Diagnostic Error Reduction. *Diagnosis.* 2016.
9. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *The American journal of medicine.* 2008;121(5 Suppl):S2-23.
10. Zwaan L, Singh H. The challenges in defining and measuring diagnostic error. *Diagnosis (Berl).* 2015;2(2):97-103.
11. Eaton DK, Kann L, Kinchen S, et al. Youth risk behavior surveillance - United States, 2011. *Morbidity and mortality weekly report Surveillance summaries.* 2012;61(4):1-162.
12. Garrison CZ, Addy CL, Jackson KL, McKeown RE, Waller JL. Major depressive disorder and dysthymia in young adolescents. *Am J Epidemiol.* 1992;135(7):792-802.
13. Whitaker A, Johnson J, Shaffer D, et al. Uncommon troubles in young people: prevalence estimates of selected psychiatric disorders in a nonreferred adolescent population. *Archives of general psychiatry.* 1990;47(5):487-496.
14. Shaffer D, Fisher P, Dulcan MK, et al. The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3): description, acceptability, prevalence rates, and performance in the MECA Study. Methods for the Epidemiology of Child and Adolescent Mental Disorders Study. *Journal of the American Academy of Child and Adolescent Psychiatry.* 1996;35(7):865-877.
15. Lewinsohn PM, Hops H, Roberts RE, Seeley JR, Andrews JA. Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-III-R disorders in high school students. *Journal of abnormal psychology.* 1993;102(1):133-144.
16. Hedden SL, Kennet J, Lipari R, et al. *Behavioral health trends in the United States: Results from the 2014 National Survey on Drug Use and Health.* Center for Behavioral Health Statistics and Quality;2015.
17. Glazebrook C, Hollis C, Heussler H, Goodman R, Coates L. Detecting emotional and behavioural problems in paediatric clinics. *Child: care, health and development.* 2003;29(2):141-149.
18. Lurbe E, Alvarez J, Redon J. Diagnosis and treatment of hypertension in children. *Current hypertension reports.* 2010;12(6):480-486.

19. Obarzanek E, Wu CO, Cutler JA, Kavey RE, Pearson GD, Daniels SR. Prevalence and incidence of hypertension in adolescent girls. *The Journal of pediatrics.* 2010;157(3):461-467, 467 e461-465.
20. Falkner B, Gidding SS, Portman R, Rosner B. Blood pressure variability and classification of prehypertension and hypertension in adolescence. *Pediatrics.* 2008;122(2):238-242.
21. Redwine KM, Acosta AA, Poffenbarger T, Portman RJ, Samuels J. Development of hypertension in adolescents with pre-hypertension. *The Journal of pediatrics.* 2012;160(1):98-103.
22. Brady TM, Neu AM, Miller ER, 3rd, Appel LJ, Siberry GK, Solomon BS. Real-time electronic medical record alerts increase high blood pressure recognition in children. *Clin Pediatr (Phila).* 2015;54(7):667-675.
23. Eder M, Smith SG, Cappleman J, Hickner J, Elder N, Singh G. *Improving Your Office Testing Process, A Toolkit for Rapid-Cycle Patient Safety and Quality Improvement.* Rockville, MD: Agency for Healthcare Research and Quality; August 2013.
24. Poon EG, Gandhi TK, Sequist TD, Murff HJ, Karson AS, Bates DW. "I wish I had seen this test result earlier!": Dissatisfaction with test result management systems in primary care. *Archives of internal medicine.* 2004;164(20):2223-2228.
25. Wahls TL, Cram PM. The frequency of missed test results and associated treatment delays in a highly computerized health system. *BMC family practice.* 2007;8:32.
26. Miller MR, Niedner MF, Huskins WC, et al. Reducing PICU central line-associated bloodstream infections: 3-year results. *Pediatrics.* 2011;128(5):e1077-1083.
27. Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *The New England journal of medicine.* 2006;355(26):2725-2732.
28. Nadeem E, Olin SS, Hill LC, Hoagwood KE, Horwitz SM. Understanding the components of quality improvement collaboratives: a systematic literature review. *The Milbank quarterly.* 2013;91(2):354-394.
29. Hulscher ME, Schouten LM, Grol RP, Buchan H. Determinants of success of quality improvement collaboratives: what does the literature show? *BMJ quality & safety.* 2013;22(1):19-31.
30. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Grol RP. Evidence for the impact of quality improvement collaboratives: systematic review. *Bmj.* 2008;336(7659):1491-1494.
31. Beers LS, Godoy L, John T, et al. Mental Health Screening Quality Improvement Learning Collaborative in Pediatric Primary Care. *Pediatrics.* 2017;140(6).
32. Greevy R, Lu B, Silber JH, Rosenbaum P. Optimal multivariate matching before randomization. *Biostatistics.* 2004;5(2):263-275.
33. Rinke ML, Singh H, Heo M, et al. Diagnostic Errors in Primary Care Pediatrics: Project RedDE. *Acad Pediatr.* 2018;18(2):220-227.
34. Rinke ML, Chen AR, Bundy DG, et al. Implementation of a central line maintenance care bundle in hospitalized pediatric oncology patients. *Pediatrics.* 2012;130(4):e996-e1004.
35. Rinke ML, Bundy DG, Chen AR, et al. Central line maintenance bundles and CLABSIs in ambulatory oncology patients. *Pediatrics.* 2013;132(5):e1403-1412.
36. Rinke ML. Toolkit for Reducing Diagnostic Errors in Primary Care - Project RedDE! Published 2018. Accessed November 15, 2018.
37. Falkner B, Daniels SR. Summary of the Fourth Report on the Diagnosis, Evaluation, and Treatment of High Blood Pressure in Children and Adolescents. *Hypertension.* 2004;44(4):387-388.

38. Zuckerbrot RA, Jensen PS. Improving recognition of adolescent depression in primary care. *Archives of pediatrics & adolescent medicine.* 2006;160(7):694-704.
39. Mayne SL, Ross ME, Song L, et al. Variations in Mental Health Diagnosis and Prescribing Across Pediatric Primary Care Practices. *Pediatrics.* 2016;137(5).
40. Sadek N, Bona J. Subsyndromal symptomatic depression: a new concept. *Depress Anxiety.* 2000;12(1):30-39.
41. Geoffrey RS, Cynthia B, Graham AB, 3rd, et al. 2014 recommendations for pediatric preventive health care. *Pediatrics.* 2014;133(3):568-570.
42. American Academy of Pediatrics Committee on Environmental H. Lead exposure in children: prevention, detection, and management. *Pediatrics.* 2005;116(4):1036-1046.
43. Dahlberg RL. *Preventing childhood lead poinsoning in NewJersey: Advocates and state government working together to increase the lead screening of children.* October 2005,
44. Nusbaum MR, Wallace RR, Slatt LM, Kondrad EC. Sexually transmitted infections and increased risk of co-infection with human immunodeficiency virus. *The Journal of the American Osteopathic Association.* 2004;104(12):527-535.
45. Ku L, St Louis M, Farshy C, et al. Risk behaviors, medical care, and chlamydial infection among young men in the United States. *American journal of public health.* 2002;92(7):1140-1143.
46. Ginocchio RH, Veenstra DL, Connell FA, Marrazzo JM. The clinical and economic consequences of screening young men for genital chlamydial infection. *Sexually transmitted diseases.* 2003;30(2):99-106.
47. Barash J. Group A streptococcal throat infection - to treat or not to treat? *Acta paediatrica.* 2009;98(3):434-436.
48. Piccardi C, Detollenaere J, Vanden Bussche P, Willems S. Social disparities in patient safety in primary care: a systematic review. *Int J Equity Health.* 2018;17(1):114.
49. Expert panel on integrated guidelines for cardiovascular health and risk reduction in children and adolescents: summary report. *Pediatrics.* 2011;128 Suppl 5:S213-256.
50. Lloyd-Jones DM, Hong Y, Labarthe D, et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation.* 2010;121(4):586-613.
51. Benedetti R, Flock B, Pedersen S, Ahern M. Improved clinical outcomes for fee-for-service physician practices participating in a diabetes care collaborative. *Jt Comm J Qual Saf.* 2004;30(4):187-194.
52. Rinke ML, Singh H, Heo M, et al. Diagnostic Errors in Primary Care Pediatrics: Project RedDE. *Acad Pediatr.* 2017.
53. Sittig DF, Singh H. Improving test result follow-up through electronic health records requires more than just an alert. *Journal of general internal medicine.* 2012;27(10):1235-1237.
54. Nolan T, Resar R, Haraden C, Griffin FA. *Improving the Reliability of Health Care.* Cambridge, MA: Institute for Healthcare Improvement;2004.