Title: Reducing Diagnostic Errors in Primary Care Pediatrics (Project RedeDE)

Research Team: Michael L. Rinke, MD, PhD, David G. Bundy, MD, MPH, Steven J. Choi, MD, Jason Adelman, MD, Heather C. O'Donnell, MD, Moonseong Heo, PhD, Keri Thiessen, MEd, Elizabeth Rice-Conboy, Hardeep Singh, MD, MPH and the American Academy of Pediatrics' Quality Improvement Innovation Network

NCT Number: 02798354

1. Background/Significance:

Despite being identified in 1999 in '*To Err is Human*' as a central facet of patient safety,[1] diagnostic errors (missed, delayed or incorrect diagnoses) remain understudied,[2] cause appreciable patient harm,[3] and their reduction lags behind advances in other patient safety areas.[4,5]. Limited pediatric-focused research on diagnostic errors highlights the significance of this problem in children: 54% of pediatricians report making diagnostic errors at least monthly and 45% report making harmful diagnostic errors at least annually.[6] In ambulatory pediatrics, evidence of harmful diagnostic errors is emerging: children experience a median delay of 3 years before receiving an asthma diagnosis,[7] and 23% of children diagnosed with retinoblastoma have a diagnostic delay of more than 8 weeks, leaving them at increased risk of tumor invasion.[8] Errors in diagnosis were the most prevalent error type in closed pediatric malpractice claims.[9] Evidence from adult and pediatric primary care settings suggest diagnostic errors are harmful and involve a variety of diseases.[5]

Few studies rigorously investigate diagnostic error reduction efforts and even fewer focus on children.[5] A recent review led by a co-investigator of this proposal, suggests a need for empirical studies to test interventions to reduce diagnostic errors, and that most existing studies focus on adults, lack rigorous process and outcome measures, suffer from non-experimental or quasi experimental designs, and/or were limited to a single institution.[10] Evidence for reducing pediatric diagnostic errors is rare and often not current: the implementation of a pediatric trauma team reduced delayed and missed diagnoses of major injuries from 4.3% to 0.46%.[11] Computer aided differential diagnoses reduced mean time to diagnosis of all pediatric inpatients from 2.8 days to 1.9 days in 1975.[12] Research on interventions to prevent pediatric diagnostic errors is needed.

The proposal will focus on 3 specific, high-risk, pediatric ambulatory diagnostic errors each representing a unique dimension of diagnostic assessment: evaluation of symptoms, evaluation of signs and follow-up of diagnostic tests. Adolescent depression (i.e. symptoms) affects nearly 10% of teenagers,[13-16] is misdiagnosed in almost 75% of adolescents[17] and causes significant morbidity.[18] Pediatric elevated blood pressure (signs) is misdiagnosed in 74-87% of patients,[19,20] often due to inaccurate application of blood pressure parameters that change based on age, gender and height. Actionable pediatric laboratory values (diagnostic tests) are potentially delayed up to 26% of the time in preliminary investigations and 7-65% in adults,[21,22] leading to harm and malpractice claims.[22-25]

We propose to conduct a multisite, prospective, stepped wedge cluster randomized trial testing a quality improvement collaborative (QIC) intervention within the American Academy of Pediatrics' Quality Improvement Innovation Networks (QuIIN) to reduce the incidence of pediatric primary care diagnostic errors. QuIIN is a national network of over 300 primary care practices, ranging from tertiary care academic medical centers to single practitioner private practices, interested in and experienced with QICs. Because many processes are likely to be common across diagnostic errors in outpatient settings, a multifaceted intervention, such as a QIC, has a high likelihood of success and broad applicability across populations. Preparatory inquiries to QuIIN primary care providers suggest high interest in reducing these 3 diagnostic errors and provider agreement with randomization to evaluate diagnostic error interventions.

<u>The purpose of this IRB application is to cover the data analysis of the QIC data generated in this project. The AAP's IRB will cover each individual pediatric clinic, the data collection and the conduct of the study.</u>

2. Study Design:

Objectives:

Primary

- To determine whether a QIC consisting of evidence-based best-practice methodologies, mini-root cause analyses, data sharing, and behavior change techniques, is associated with a reduction in 3 specific diagnostic error rates in a national group of pediatric primary care practices.
    - *Hypothesis 1: Implementation of a QIC will lead to a 40% reduction in missed diagnosis of adolescent depression.*
    - *Hypothesis 2: Implementation of a QIC will lead to a 30% reduction in missed diagnosis of pediatric elevated blood pressure.*
    - *Hypothesis 3: Implementation of a QIC will lead to a 45% reduction in delayed diagnosis of actionable laboratory results.*

Secondary

- To determine if a QIC's effect changes for wave 1 versus wave 2 participants, who serve as the control group in the first year of the collaborative.
- To further investigate the epidemiology of three ambulatory pediatric diagnostic errors: missed diagnosis of adolescent depression, missed diagnosis of pediatric elevated blood pressure, and delayed diagnosis of actionable laboratory results.
- To evaluate patient outcomes related to these diagnoses including outcomes after positive depression screening, missed elevated blood pressure screening and delayed diagnosis of actionable laboratory values.

Methods:

This proposal aims to identify systems solutions for reducing harmful pediatric diagnostic errors. We will perform a prospective, stepped wedge cluster randomized controlled trial to test the hypothesis that QIC strategies can reduce three specific diagnostic errors in a national group of ambulatory pediatric primary care practices. These data will inform future diagnostic error prevention efforts and provide data on QIC efforts aimed at pediatric diagnostic errors. Baseline data from the QIC will identify incidence data for errors in 3 separate stages of the diagnostic process: symptoms (missed diagnosis of adolescent depression), signs (missed diagnosis of pediatric elevated blood pressure) and diagnostic tests (delayed diagnosis of actionable laboratory results). Each of these diagnostic errors represents missed opportunities to recognize abnormal diagnostic findings.[26-28] Human subjects' protection will be completed at the CHAM to cover the researchers and data analysis and at the American Academy of Pediatrics to cover all practices, as completed for prior QuIIN projects.[29]

The goal of this proposal is to reduce diagnostic errors by tracking diagnostic errors rates (outcomes measures) and reliably performing best practices for diagnosing adolescent depression, pediatric elevated blood pressure and actionable laboratory results (process measures). In order to reduce overall data collection burden on practices and to increase engagement of practices that would have been in the control group, we have changed the study design. These changes do not affect the statistical power of the study. The intervention will consist of implementing QIC methodology in all practices to assist teams in reliably performing the processes that will reduce the diagnostic error outcomes via evidence-based tools.[30-32] Practices will be randomized into three groups, with each group intervening on 1 diagnostic error and serving as the control group for a second diagnostic error. After 8 months of participation, each group will begin intervening on the diagnostic error they were a control group for, continue intervening on their original diagnostic error, and function as the control group for the third diagnostic error. In the final 8 months of the QIC, all three groups will intervene on all three diagnostic errors.

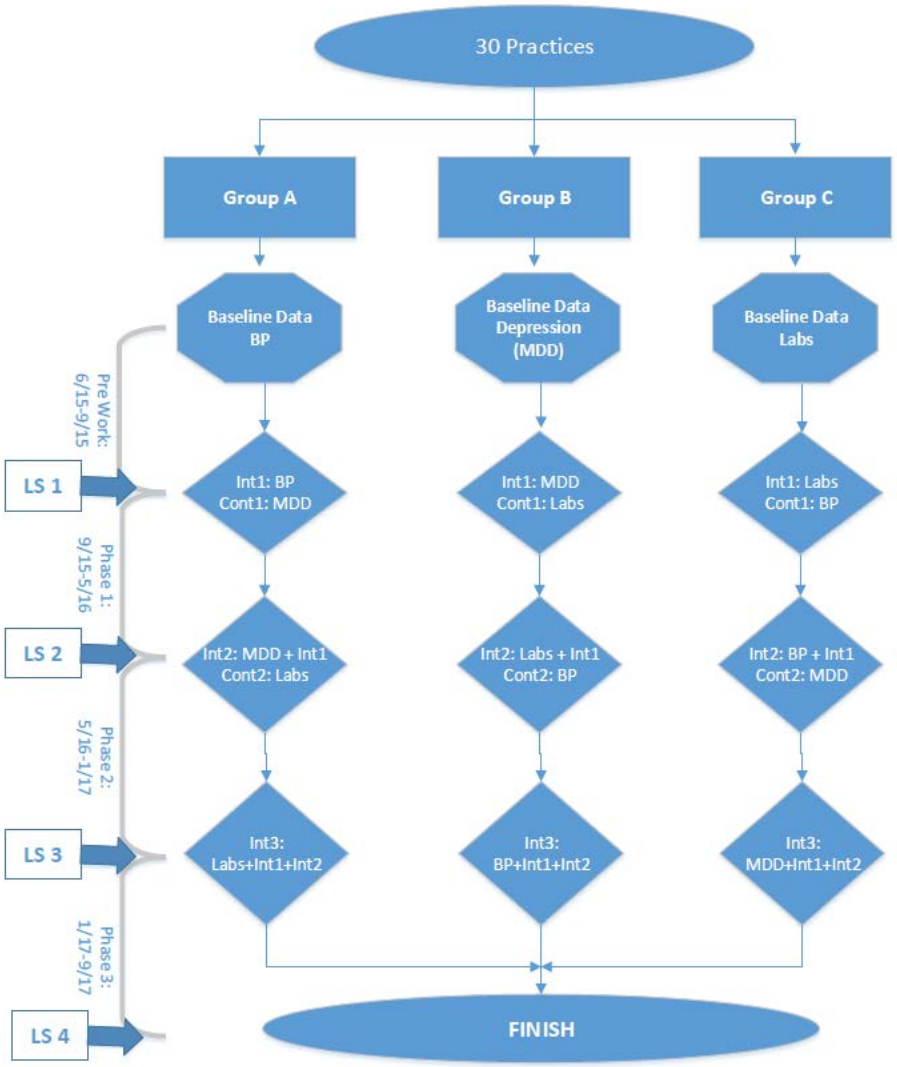


Table 1: Quality Improvement Collaborative (QIC) Components

- Bi-annual 2 day interactive webinar learning sessions
- Monthly all collaborative webinar & conference calls sharing best practices
- Monthly one on one team calls with dedicated QI coach (QuIIN staff and research team)
- Monthly data submission on both process and outcome measures (diagnostic errors) using QuIIN IT infrastructure
- Monthly data feedback both at aggregate level with full inter-team transparency as well as at specific institutional level using QuIIN IT infrastructure
- Monthly mini-root cause analyses performed on 3 errors at each site
- Multidisciplinary teams consisting of at least a physician, nurse and office practice associate
- Instruction on best practices from content area experts in QI, diagnostic errors, hypertension, mental health and EHRs
- Instruction on model for improvement, MUSIQ and behavior change via QI methodology (small tests of change/PDSA
- Instruction on QI team leadership and team QI skills
- Ongoing sharing of best ideas and barriers/issues among institutional teams

In order to test the hypothesis that a QIC can reduce these diagnostic errors, we will employ a prospective, stepped wedge cluster randomized controlled trial. First, all practices will submit baseline data on one of the three diagnostic errors (defined below) and this will provide data for an initial manuscript on the epidemiology of these diagnostic errors. Following baseline data collection, we will apply multivariate matching before randomization, which has been shown to provide more accurate effect estimates, creating an

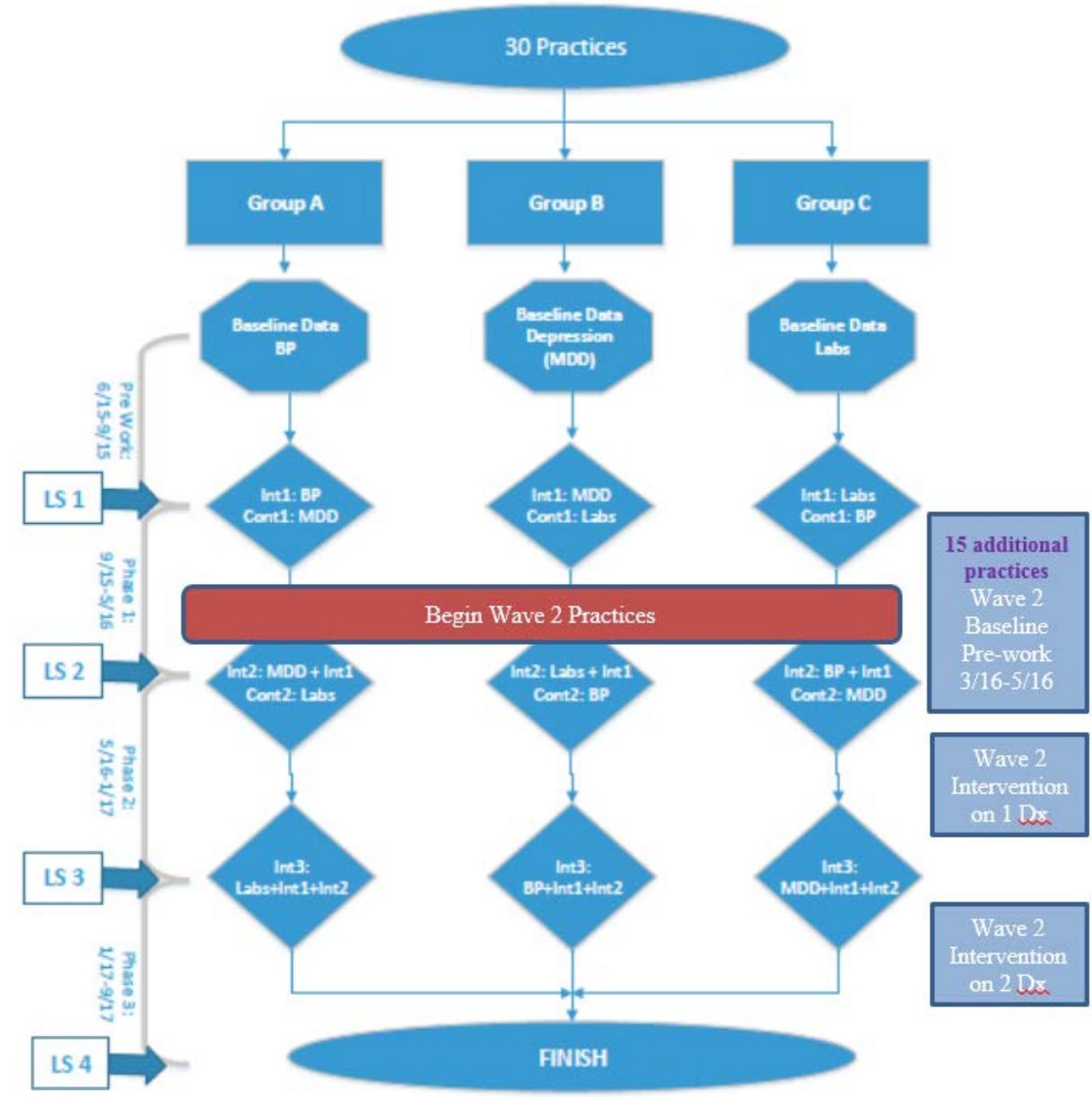
equivalent of 7% increase in sample size.[33] Practices will be matched based on key demographics related to diagnostic errors including 1) patient volume per practitioner +/- 100 patients, 2) university affiliation and 3) prior work reducing these diagnostic errors. After matching, randomization will occur via computerized random number generation and one-third of practices will be assigned to each of the three QIC intervention groups.

The QIC intervention components are described in Table 1 and closely mirror prior evidence on components integral to QIC success[32,34,35] and prior QICs conducted by QuIIN and the research team.[29,36] Prior research by co-investigators and others have shown that reducing diagnostic errors requires more than just technological alerts,[37,38] which is why a QIC intervention was chosen. The QIC will ensure the 6 key contextual MUSIQ factors will be met,[39] increasing the ability of each practice to reduce these errors. Rapid data feedback on performance with benchmarking is crucial to behavior change for diagnostic error reduction.[40] Root cause analyses have proven useful in identifying underlying causes of diagnostic errors.[41] The mini-root cause analyses used in this QIC involve identifying 1 diagnostic error monthly at each QIC site, examining standardized patient, provider, and systems factors that could have led to the error, and spreading systematic lessons learned to prevent future errors.[42,43] These factors could include 1) patient factors of age, gender, reason for visit, patient comorbidities, language barriers, acute illness, agitation of patient/family, social issues, etc. , 2) provider factors of type of provider, provider level of training, provider fatigue/impairment, personal stressors of providers, provider disagreements, provider knowledge, provider beliefs about the project or the patient, etc. and 3) systems factors of time of visit, clinic milieu during visit (chaotic vs. calm), increased workload, staffing concerns, verbal/written communication, computer software or hardware, non-computer equipment, etc.. These deidentifed factors will be submitted to the QIC monthly in order to aggregate causes of these errors. No protected health information (PHI) will be collected in the aggregate entered data for these min-root cause analyses. One of the novel approaches of this collaborative is the use of an interactive webinar format for learning sessions.

All practices will participate in the QIC for the full 2 years of the study. They will attend an initial 2-day interactive webinar learning session where they will learn QI methodology, enhance and practice QI skills, identify local 30-60 day aims for the diagnostic error they are intervening on first, increase their understanding of diagnostic errors and these errors in particular, and create systems for standardized data collection and data entry. Practices will submit monthly data on their intervention and control diagnostic errors but only receive centralized data feedback, coaching or collaborative involvement regarding their intervention error. (Figure 1) We acknowledge that error measurement in the baseline and intervention periods may introduce an intervention effect, but we have no reason to believe this effect will be different for the intervention and control practices. As noted above, after 8 months, practices will participate in a second 2-day video conference learning session, continue to intervene on their first error, begin to intervene on the error they were a control group for and begin to collect control data for the third diagnostic error. Including the control group in the QIC ensures that harm is prevented in the maximum number of patients and that clinical equipoise is met, as we believe this intervention will reduce errors. This stepped wedge design with two steps allows for the primary outcome comparison (diagnostic errors in intervention versus control) and secondary outcome comparisons: 1) diagnostic error rates in second wave participants), 2) Sustainability after first eight months and 3) collaborative effect for second wave practices. These secondary outcome comparisons will shed light on the effectiveness and sustainability of QICs and inform future collaborative designs.

Given concerns about practice attrition, we will additionally recruit a second wave of practices to join the collaborative. The three phase quality improvement research model lends itself to simply on-boarding a new group of practices and integrating them into the project in Phase 2 which would begin in June of 2016. The up to 15 Wave 2 practices would be recruited

in the winter of 2015-2016, similarly randomized to the three already established collaborative groups and enrolled in March of 2016, collect baseline data comparable to the wave 1 practices and then participate in the collaborative beginning in June of 2016. As noted above, after 8 months, practices will participate in a second 2-day video conference learning session, continue to intervene on their first error, begin to intervene on the error they were a control group for and begin to collect control data for the third diagnostic error. The only difference between the Wave 1 and Wave 2 practices, are that the Wave 1 practices will ultimately intervene on all three diagnostic errors of interest. As the Wave 2 practices are entering the collaborative 8 months later, they will only intervene on two of the three diagnostic errors of interest over the course of 16 months. At the end of the collaborative, all resources and tools for the third diagnostic error will be given to the Wave 2 participants.



Target Diagnostic Error Outcome Measures:

All outcome metrics are described in detail in Table 2 and the Data Collection section below.

The missed diagnosis of adolescent depression primary outcome measure is defined as the incidence of depression in adolescents ≥11 years old compared in intervention versus control practices. This outcome measure was chosen because prior research suggested up to 75% of adolescent cases are missed by primary providers,[17,44] allowing incidence to serve as a proxy for missed diagnoses. The research team strongly considered using documented depressive symptoms (e.g. poor school performance, interrupted sleep patterns, increased disruptive behaviors, etc.) without appropriate provider identification or referral to mental health evaluation as the missed diagnosis of adolescent depression primary outcome. This metric was not chosen because preliminary data suggested the number of adolescents with symptoms suggestive of depression documented prior to a mental health referral was extremely small. This finding echoes research that pediatricians' use of adolescent and parental chief complaints to identify depressive risk factors consistently under identifies adolescent depression.[44,45] The process measure for this diagnostic error will be the percent of adolescents screened for depression, a practice recommended for all adolescents by the USPSTF,[46] and feasible and acceptable to primary care pediatricians.[47-49]

Table 2: Outcome and Process Measures, and Evidence-Based Tools for Diagnostic Error Reduction

Primary Outcome Measure	Process Measures	Evidence-Based Tools
Number of Adolescents Diagnosed with Depression per 100 Adolescent Visits Patients ≥11 years old with documentation of major depression or subsyndromal depression diagnoses	1. Number of adolescents screened for depression with the PHQ-9-M per total number of adolescents seen per month	1. Patient screening in the waiting room 2. Triage nurse scoring and flagging to facilitate provider recognition
Elevated Blood Pressure Not Recognized by Providers per 100 Elevated Blood Pressures Measured Systolic or Diastolic Blood Pressure ≥ 90th percentile for age, gender and height or ≥120/80 in ≥3 years old patients without history of hypertension and none of: 1) provider repeated blood pressure, 2) clinic note mentions elevated blood pressure/hypertension 3) plan included recheck or evaluation of blood pressure, or 4) ordering laboratory or other studies to evaluate elevated blood pressure	1. Number of patients ≥3 years old with blood pressure checked at triage per total number of patients ≥3 years old seen	1. Triage nurse or EHR flagging of elevated blood pressures to facilitate provider recognition 2. Family notification of elevated blood pressure at triage
Delayed Actions on Laboratory Results per 100 Abnormal Laboratory Results Received[50] No documented action step for first positive after 30 days: 1. Hgb≤11 and MCV≤75 in 1 or 2 year old→ documentation of beginning iron, sending iron studies or family conversation 2. Lead≥5→documentation of family conversation on lead remediation or plan to retest No documented action step for first positive after 7 days: 1. Positive Gonorrhea, Chlamydia, Syphilis or HIV test→ documentation of antibiotics begun or referral to HIV specialist 2. Positive group A *streptococcal* throat culture with negative rapid group A *streptococcal*	1. Percent of providers with unread or unacknowledged laboratory results in their EHR inbox for greater than 72 hours	1. Weekly nurse review of all abnormal laboratory results 2. Standardized protocols for handling laboratory results 3. Standing orders for laboratory results action steps to be completed by nurse with appropriate provider oversight

test→documentation of antibiotics begun or family conversation 3. TSH≤0.5 or ≥4.5 in ≥1 year old→plan to repeat lab values or referral to endocrinologist		

Early detection via screening represents the best opportunity to reduce missed diagnosis of adolescent depression. We will screen patients with the Patient Health Questionnaire 9-item Modified for Adolescents (PHQ-9-M), as it is a validated and highly specific tool (94.7%).[51] Building on work done by Ruth E. Stein, MD,[52,53] a member of the Project Expert Group and others,[49] we will use coaching and learning sessions to ensure appropriate follow-up for identified mental health concerns are in place before practices begin universal screening. The evidence-based tools utilized by the collaborative to improve diagnosis of adolescent depression include patient screening in the waiting room, followed by nurse scoring and flagging. This methodology increased adolescent screening rates to 80-84% in QI interventions at CHAM and in 6 non-affiliated sites,[49] and is a necessary first step to adolescent depression diagnosis.

The missed diagnosis of pediatric elevated blood pressure primary outcome is defined as elevated systolic or diastolic blood pressure for age, gender and height without provider recognition per 100 elevated blood pressures measured in intervention versus control practices. The process measure will assess the percent of children ≥3 years old receiving blood pressure measurement at triage, as this is standard of care[54,55] and an endorsed recommendation by the AAP,[56] but not regularly adhered to in all pediatric clinics.[57,58] The evidence-based tools implemented by the collaborative to help reduce this diagnostic error will involve triage nurse or EHR flagging of elevated blood pressures for provider recognition[19] and family notification of elevated blood pressures.

The delayed diagnosis of actionable laboratory results primary outcome measure is defined as the number of delayed actions, as defined in Table 2, per 100 abnormal laboratory results received in intervention versus control practices. Abnormal values are defined by pediatric reference texts.[50] Since no standard definition of an actionable laboratory 'delay' currently exists, timeliness of follow-up was determined by expert opinion and discussions with primary care practitioners. Delays are defined as greater than 30 days for action on elevated lead levels or microcytic anemia, or greater than 7 days for action on STI treatment, *streptococcal* pharyngitis treatment or elevated or reduced TSH work-up. The process measure for delayed diagnosis of actionable laboratory results will be the number of providers with laboratory results unread or unacknowledged in their EHR inbox for more than 72 hours. The tools implemented by the collaborative will include standing orders for nurse actions when abnormal test results are received. For example, if a patient has a positive Chlamydia test, the nurse will immediately call the patient and schedule the next available visit for counseling and antibiotics. These interventions are based on successful newborn screening QICs at CHAM and QuIIN,[29] and AHRQ's toolkit for improving office-based testing.[59]

QIC Data Collection and Standardization:

An initial orientation session will teach all practices to identify each of these 3 diagnostic errors via chart review, collect data and reliably enter it into the QuIIN IT system, QIDA. The QIDA system has been used by prior QuIIN projects, is web-based and HIPAA compliant, and QuIIN will take responsibility for data security. QuIIN's project manager has experiencing ensuring accurate and timely data submission for primary care pediatric practices from prior QuIIN QICs.

Table 3: QIC Primary Outcome Data Collection for Each Practice

Baseline (One Time Data Collection)	**QIC Intervention Period** (Monthly Data Collection)
Missed Adolescent Depression	
50 ≥11 year old patient charts who visited clinic in prior 6 months. Chart screened for depression diagnosis.	17 ≥11 year old patient charts who visited clinic in prior month. Chart screened for depression diagnosis.
Missed Elevated Blood Pressure	
50 ≥ 3 year old patient charts with elevated blood pressure recorded at triage. Chart screened for provider recognition.	10 ≥ 3 year old patient charts with elevated blood pressure recorded at triage. Chart screened for provider recognition.
Delayed Diagnosis of Actionable Laboratory Results	
50 patient charts with abnormal laboratory results as defined above. Chart screened for appropriate action within time guidelines defined above.	10 patient charts with abnormal laboratory results as defined above. Chart screened for appropriate action within time guidelines defined above.

*All charts screened will have the following demographic data recorded: age, gender and insurance status.

Practices will collect the primary outcome data by employing EHR based record review or paper-based methods depending on abilities and interest of local EHRs. Developing data collection tools for both EHR and paper based practices allows for maximum generalizability of lessons learned from this collaborative. These tools will ensure standardized data collection by disparate practices and were pilot tested with two QuIIN practices. All data collection tools will be pilot tested on paper and beta tested by a minimum of 2 Expert Group members (always including the QI Advisor). As described below, data will be checked for outliers or abnormal values and feedback given to practices to ensure consistency across practices. Additional data on providers' abilities to recognize abnormal results, either in mental health, blood pressure or laboratory values, will also be collected as this is an important first step in the diagnostic process. These data are described further in the Data collection section below. QIC leadership will use QIDA entered data to disseminate monthly feedback reports with full transparency and QIC benchmarking to intervention practices. Consented core QI team members from each practice will be able to access their team's data. All practices will report demographic data annually, including patient volume, number and training level of physician or physician assistant-level providers, number of nursing and office staff, university-affiliation, EHR availability, presence of trainees, estimated length of time providers take with patients, and prior self-reported work aiming to improve these diagnostic errors.

Table 3 summarizes the primary outcome data each practice will collect and the types of charts they will screen. Numbers of charts are based on power calculations described below and will be a convenience sample. Adolescents with depression will be identified by monthly chart reviews or EHR data summaries when this diagnosis is mentioned in the problem list or in assessment sections of visit notes. Pediatric elevated blood pressures will be identified by flagging blood pressure measurements at triage or pulling reports from EHR systems. Similar to the depression outcome, provider recognition will be identified when this diagnosis is mentioned in the problem list or in assessment sections of visit notes. Finally, abnormal lab results will be tracked by either EHR laboratory summary reports or laboratory management registry tools, such as those found in the AHRQ toolkit for improving office-based testing,[59] and provider recognition will be identified when a diagnosis is mentioned in the problem list, in assessment

sections of visit notes or if medications are prescribed. Intervention practices will also report process outcomes (Table 2).

Finally, in order to better investigate if reducing these diagnostic errors leads to improved patient outcomes, we will ask practices one time during the 2 year collaborative to identify the time between when the appropriate information to diagnose the patient was available and when the actual diagnosis was conveyed to the family. Practices who have intervened on a given error will track 50 patient charts with one of the following metrics: 1) time between positive mental health screen and diagnosis of depression or depression diagnosis ruled out, 2) the number of days between blood pressure elevation in the medical record and notification of the family, or 3) time between abnormal laboratory result received and notification of family. Additionally, we will ask practices to evaluate 50 patients with a delayed diagnosis of actionable laboratory results 6 months after the diagnostic opportunity to evaluate for emergency department visits, clinic visits, ultimate diagnosis, clinical worsening in laboratory values and other patient morbidities. Similarly, we will ask intervention practices to evaluate 50 patients whose diagnostic instrument of the PHQ-9M was positive 6 months after the diagnostic opportunity to evaluate how many were ultimately diagnosed with depression, attended initial mental health appointments and were referred to mental health or were diagnosed directly by the pediatrician.

4. Study Population

Our cohort will consist of at least 30 primary care pediatric practices which are part of QuIIN. We are confident that we can obtain this minimum number given that 45 of the 51 QuIIN practices (88%) participating in the pre-survey were interested in participating in a diagnostic error project, and over 300 practices are part of QuIIN from 46 states. We were able to recruit and consent 34 practices, but 9 were lost after randomization due to multiple issues including change in clinic staffing, change in clinic leadership and difficulty with data collection. For this reason we will recruit 15 additional practices to join the collaborative in Wave 2. Interested practices reported seeing between 3,000 and 80,000 pediatric patients per practice annually (median 12,000). No practices will be excluded based on size as even the smallest practices have sufficient patient panels to collect the data described above. Each practice-based team will be multidisciplinary, consisting of at least one physician, nurse and office practice associate allowing for microsystem issues to be transmitted to practice leadership and QI skill to be disseminated across key groups.[39]

Inclusion Criteria

We will include 30 primary care pediatric practices that are part of the American Academy of Pediatrics' QuIIN organization. Our second wave will recruit 15 additional practices. No practices will be excluded based on race, ethnicity, gender or location.

Given that this study is minimal risk and the data submitted is deidentified, we will request a waiver of consent to look at patients' protected health information submitted by these practices. Patient Protected Health Information (PHI) will not be collected in QIDA.

Exclusion Criteria

We will not exclude any practices based on gender, race or ethnicity of its providers or patients.

4. Participant Recruitment:

Upon IRB approval, primary care clinical teams will be recruited from the membership of the AAP's Quality Improvement Innovation Networks (QuIIN) as well as other AAP listservs. In addition, Expert Group members will assist in recruiting potential practices.

Practices will be enrolled via email and phone solicitation from the 300 QuIIN practices currently involved in the organization. This procedure has been successful in 39 prior and current QuIIN quality improvement collaborative practices.[29] An application packet will be developed that will include a cover letter signed by Drs. Michael Rinke and David Bundy who represent the Project Expert Group. This application packet will also include a project overview, to describe the project and provide specific responsibilities of the practice/individual team members and a link to the application to be completed via Survey Monkey. The application will ask for basic practice information and patient and practice demographics. QuIIN members and others will be notified of the opportunity to participate via email. The application packet includes details about an informational call to be held during recruitment so any interested individual can learn more about the project before completing an application. This informational call will be conducted by Expert Group leadership, AAP QuIIN Staff, and the QI Advisor. This same recruitment format will be applied to the Wave 2 practices.

5. Informed Consent:

The core improvement team (physician, nurse and office associate health care providers) across the 30 Wave 1 practices (25 retained) and 15 Wave 2 practices, enrolled via the AAP's Quality Improvement and Innovation Networks (QuIIN) will be the primary human subjects for this proposal. Each practice will consent for participation via a standardized and previously used QuIIN consent form. Over 300 practices are part of QuIIN from 46 states and have between 1 and 80 providers in each practice. Interested practices reported seeing between 3,000 and 80,000 pediatric patients per practice per year (median 12,000). As the proposal is geared toward improving providers' actions and behaviors to be consistent with best-practice care to diagnose the 3 specific conditions outlined in this proposal, the risk to individual providers is negligible and individual consent will be covered by a waiver of consent. It would be impractical to consult every healthcare provider in every clinic in this study. This consent procedure will be carried out by the American Academy of Pediatrics and QuIIN as in previous projects conducted by their organization. Providers are typically between 30-60 years of age and no participants will be allowed who are less than 22 years old. No provider will be excluded. Similarly, because patients will experience minimal risk and only the minimal personal health information will be collected on patients, will obtain a waiver of consent for patient data. Both of these waivers have been granted on prior QuIIN projects via the AAP's IRB.

6. Risk/Benefit:

Risks:

The physician, nurse and office associate health care providers at each of the QuIIN practices will be the primary human subjects for this work. As the proposal is geared toward improving providers' actions to be consistent with best-practice care, the risk to providers is minimal. No individual provider compliance data or diagnostic error data will be recorded. Only aggregated data at the practice level is recorded. There will be no additional medical risks in implementing these procedures as they are all standard of care and we are only working to ensure reliable implementation. Providers may feel discouragement if they are unable to reliably perform certain best-practices. In order to minimize this risk, we will provide constant encouragement, utilizing systems focused, blame neutral coaching methodology. Parents and patients will be free to decline any aspect of care as occurs normally in care. As such, no additional risk is posed to the clinic patients.

There are no foreseeable financial risks to patients as a result of this study. Providers could experience a potential decrease in patient volume and therefore financial revenue if their

practice is worse than collaborative baseline and this information is made public. Practices will sign a non-disclosure agreement regarding all data, and QuIIN has never experienced a data breech in 39 prior QIC projects. The research group and QuIIN will work hard to instill this understanding of confidentiality into the collaborative. On the contrary, practices could experience an increase in revenue if patients appreciate improved care and better diagnostic accuracy in their clinics. The financial risks to providers and practices are low.

Benefits:

The potential benefits of this study to the providers include timely and actionable feedback, and alerts to help providers better recognize aspects of diagnostic performance in need of improvement. Teams will have the opportunity to test strategies for improving care that will become a national model, to work with colleagues from around the country, to learn from national experts in decreasing the incidence of diagnostic errors, utilize tools to screen for adolescent depression, high blood pressure and appropriate follow-up with actionable labs, to access practical tools and effective strategies for how to deliver better care, and receive ongoing support for improvement. If improvements in care are achieved, the benefits to parents and children are also likely to be significant. Teams will also have the opportunity to receive coaching from experts in the fields of quality improvement, medical informatics, health IT solutions, pediatric hypertension, pediatric mental health and diagnostic errors. Pending approval, pediatrician participants may have the opportunity to receive Part 4 Maintenance of Certification (MOC) credit if they meet the minimum eligibility criteria.

Although the subjects are the providers, pediatric patients may benefit from this study by being less likely to experience a diagnostic error.

7. Data Collection:

As with any study, we will take significant steps to safeguard all personal health information collected. We will only collect the minimal amount of personal health information required as described above. All files will be deindentified and disassociated from practice identifiers as soon as the study is complete. All records will be kept in locked offices, on password protected computers and in password protected files. Finally, all data submitted to the QuIIN QIDA national collaborative will be deidentified to the extent described above. QuIIN, using the QIDA software, has performed these activities on multiple prior QIC projects without any breaches of confidentiality. The QIDA system is web-based and HIPAA compliant, and QuIIN will take responsibility for data security. Only the research team will have access to the raw QIDA data, and all data will be kept on password protected computers in locked offices. No patient identifiable information is stored in the QIDA database. Project data will be stored for 7 years in the QIDA system, but once a project closes, only AAP QIDA staff will have access to the data. As such the risk of a breach of confidentiality is minimal and the legal risks associated with a breach of confidentiality are also be minimal.

There will be 6 sets of data submitted to the QuIIN Quality Improvement Data Aggregator (QIDA) program. First, all practices will report demographic data annually, including patient volume, number and training level of physician or physician assistant-level providers, number of nursing and office staff, university-affiliation, EHR availability, presence of trainees, and prior self-reported work aiming to improve these diagnostic errors. This demographic practice-level data will not include any PHI, aside from the contact information sought as with most project applications initially. The project application will serve as the first set of practice-based demographic data and it will be requested 2 additional times.

Second, each practice will collect data on 1) incidence of adolescent depression (number of adolescents diagnosed with depression or subsyndromal depression as defined by

the problem list or patient notes per 100 adolescent visits), 2) missed diagnosis of pediatric elevated blood pressure (elevated systolic or diastolic blood pressure for age, gender and height without provider recognition per 100 elevated blood pressures measured) and 3) delayed diagnosis of actionable laboratory results (number of delayed actions, as defined in Table 2, per 100 abnormal laboratory evaluations received). All of this data will originate from paper or electronic patient charts. For each of the charts screened, the following demographic data will also be recorded: age, gender and insurance status. These data will be entered into QIDA without patient names or birthdates. No PHI will be recorded as part of this dataset that will be collected monthly.

The third set of data submitted will be process measure data for each of the three diagnostic errors described in the grant. These data will be submitted by practices assigned to the related diagnostic error while they are actively intervening on that error (i.e. for 8 months the group randomized to intervene on adolescent depression will collect process measure data on screening for adolescent depression). The three process measures will be 1) the percent of adolescents screened for depression, 2) the percent of children $\geq$3 years old receiving blood pressure measurement at triage, and 3) the percent of providers with laboratory results unread or unacknowledged in their EHR inbox for more than 72 hours. These data will be entered in aggregate into QIDA for each clinic monthly and collected via electronic health record or practice based paper charting methods in each practice.

The fourth set of data collected will be mini-root cause analyses, which identify patient and systems level factors that could have contributed to 1 diagnostic error monthly. Each practice will examine standardized patient, provider and systems factors that could have led to the specific diagnostic errors identified in this grant. These factors could include 1) patient factors of age, gender, reason for visit, patient comorbidities, language barriers, acute illness, agitation of patient/family, social issues, etc. , 2) provider factors of type of provider, provider level of training, provider fatigue/impairment, personal stressors of providers, provider disagreements, provider knowledge, provider beliefs about the project or the patient, etc. and 3) systems factors of time of visit, clinic milieu during visit (chaotic vs. calm), increased workload, staffing concerns, verbal/written communication, computer software or hardware, non-computer equipment, etc.. These mini-root cause analyses factors will be collected on paper based forms, deidentified and entered into SurveyMonkey. Practices will be encouraged to identify and spread lessons learned to all providers in their practices to prevent future errors.

The fifth set of data collected will be related to providers' abilities to recognize abnormal findings that could relate to the diagnostic errors of interest. As recognition of abnormal findings is key to ultimately making a correct diagnosis, it is important for practices to track and understand if their clinicians are appropriately recognizing abnormal results. These data will be obtained by electronic and paper chart review of the same 10-17 charts per month as the primary outcome measures noted above. They will include: 1) Percent of charts with documentation of patient's systolic and diastolic blood pressure percentiles documented in the record, 2) Percent of charts with documentation of mental health concerns or exclusion of mental health concerns in patient record or 3) Percent of charts with abnormal actionable laboratory results noted as "abnormal" in the patient record. Clinics will only collect these data when they are the intervention or control group for a given measure.

The final set of data will be aggregate data of patients one time during the 2 year collaborative to identify the time between when the appropriate information to diagnose the patient was available and when the actual diagnosis was conveyed to the family. Practices who have intervened on a given error will track 50 patient charts with one of the following metrics: 1) time between positive mental health screen and diagnosis of depression or depression diagnosis ruled out, 2) the number of days between blood pressure elevation in the medical record and notification of the family, or 3) time between abnormal laboratory result received and notification of family., Additionally, we will ask practices to identify, in the same 50 patients with

a delayed diagnosis of actionable laboratory results 6 months after the diagnostic opportunity to evaluate for emergency department visits, clinic visits, ultimate diagnosis, clinical worsening in laboratory values and other patient morbidities. Similarly, we will ask intervention practices to evaluate 50 patients whose diagnostic instrument of the PHQ-9M was positive 6 months after the diagnostic opportunity to evaluate how many were ultimately diagnosed with depression, attended initial mental health appointments and were referred to mental health or were diagnosed directly by the pediatrician.. This final set of data will be collected only once between 12-18 months after the intervention period begins.

For research and publications that may result from this work, all data will be reported in aggregate, and individual and practice data will not be identifiable. If practice data is presented, each practice will receive an ID number in the report. Potential publications may include a conceptual model of key barriers and potentially useful strategies that emerged from this project. No patients or practice staff will be identified in any report or publication about this study. Practice names will only be used in the acknowledgement section of any potential publication.

Statistical Models: The primary independent variable will be assignment to the control or QIC group and the 3 primary dependent variables will be the adolescent depression rate (number of adolescents diagnosed with depression per 100 adolescent visits), the missed pediatric elevated blood pressure rate (elevated blood pressure not recognized by providers per 100 elevated blood pressures measured) and the delayed diagnosis of actionable laboratory results rate (delayed actions on laboratory results per 100 abnormal laboratory results received). Since the individual-level outcome will be binary (missed/delayed diagnosis or not), we will apply mixed-effects logistic models for testing the three primary hypotheses during the first study year and for testing secondary hypotheses. The clinic-specific intercept will be considered random to take into account potential correlations of outcomes from the subjects within the clinics. We will test each of the three main effects using a mixed-effects model. All primary and secondary analyses will be controlled for potential confounding effects including factors such as baseline diagnostic error rates, patient volume, number and training of physician or physician assistant-level providers, number of nursing and office staff, university-affiliation, EHR availability, presence of trainees, and prior work improving these diagnostic errors.

Power Analysis: Assuming a two sided alpha level of 0.05 and >80% power, we calculated the effect size we could detect for each of the 3 diagnostic error primary outcome measures, comparing the 10-20 intervention practices, given the stepped wedge design, to the 10-20 control practices using the above models, and the QIC methodology and preliminary data. 1) Missed diagnosis of depression: If each practice screens 17 adolescent charts a month, we are able to detect an effect size of 2.4% or a reduction in the diagnostic error rate as small as 33% (intra-class correlation coefficient (ICC)=0.005). 2) Missed diagnosis of pediatric elevated blood pressure: If each practice screens 10 patient charts with elevated blood pressure a month, we are able to detect an effect size of 12% or a reduction in the diagnostic error rate as small as 13% (ICC=0.1). 3) Delayed diagnosis of actionable laboratory results: If each practice collects 10 patient charts with abnormal laboratory values a month, we are able to detect an effect size of 13%, or a reduction in the diagnostic error rate as small as 15% (ICC=0.1). Although we are testing three outcomes, we do not correct the alpha level for multiple testing since the study subjects will most likely be different across the three outcome comparisons.

8. Data Quality Control and Database Management:

We will make every effort to minimize data recording errors by checking for outliers or abnormal values using frequency tables or graphical tools such as box plots. If noted, we will

correct or check the validity of those values and discuss with practices. Although we expect few by virtue of randomization, we will examine imbalances between groups with respect to baseline practice demographic characteristics and outcome measurements. We will use standard statistical methods such as chi-square, Fisher exact, two sample t-tests or Mann-Whitney tests, depending on the outcome scale and validity of normal distribution assumptions. If necessary, we will apply log-transformations for right-skewed data. We will include variables that are significantly different in the above statistical models to control for their potential confounding effects. Although we anticipate little missing data based on prior QICs, when noted, we will examine reasons for incomplete records. We will analyze available data as the primary analytic strategy since mixed-effects modeling is valid for missing at random data. We will also conduct sensitivity analyses applying a full conditional specification multiple imputation method which can apply to missing not at random data. We will compare results between available data analysis and multiple imputations data analysis.

References:

1. Institute of Medicine. To err is human: Building a safer health system. Washington, D.C.: National Academy Press; 1999.
2. Zwaan L, Schiff GD, Singh H. Advancing the research agenda for diagnostic error reduction. BMJ quality & safety 2013;22 Suppl 2:ii52-ii7.
3. Graber ML. The incidence of diagnostic error in medicine. BMJ quality & safety 2013;22 Suppl 2:ii21-ii7.
4. Singh H, Naik AD, Rao R, Petersen LA. Reducing diagnostic errors through effective communication: harnessing the power of information technology. Journal of general internal medicine 2008;23:489-94.
5. McDonald KM, Matesic B, Contopoulos-Ioannidis DG, et al. Patient safety strategies targeted at diagnostic errors: a systematic review. Ann Intern Med 2013;158:381-9.
6. Singh H, Thomas EJ, Wilson L, et al. Errors of diagnosis in pediatric practice: a multisite survey. Pediatrics 2010;126:70-9.
7. Charlton I, Jones K, Bain J. Delay in diagnosis of childhood asthma and its influence on respiratory consultation rates. Arch Dis Child 1991;66:633-5.
8. Goddard AG, Kingston JE, Hungerford JL. Delay in diagnosis of retinoblastoma: risk factors and treatment outcome. The British journal of ophthalmology 1999;83:1320-3.
9. Carroll AE, Buddenbaum JL. Malpractice claims involving pediatricians: epidemiology and etiology. Pediatrics 2007;120:10-7.
10. Singh H, Graber ML, Kissam SM, et al. System-related interventions to reduce diagnostic errors: a narrative review. BMJ quality & safety 2012;21:160-70.
11. Perno JF, Schunk JE, Hansen KW, Furnival RA. Significant reduction in delayed diagnosis of injury with implementation of a pediatric trauma service. Pediatric emergency care 2005;21:367-71.
12. Wexler JR, Swender PT, Tunnessen WW, Jr., Oski FA. Impact of a system of computer-assisted diagnosis. Initial evaluation of the hospitalized patient. American journal of diseases of children 1975;129:203-5.
13. Garrison CZ, Addy CL, Jackson KL, McKeown RE, Waller JL. Major depressive disorder and dysthymia in young adolescents. Am J Epidemiol 1992;135:792-802.
14. 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;47:487-96.
15. 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;35:865-77.
16. 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;102:133-44.
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;29:141-9.
18. Eaton DK, Kann L, Kinchen S, et al. Youth risk behavior surveillance - United States, 2011. Morbidity and mortality weekly report Surveillance summaries 2012;61:1-162.
19. Brady TM, Neu AM, Siberry G, Solomon B. Increased Provider Recognition of Elevated Blood Presure in Children. American Society of Nephrology; 2012; San Diego, CA.
20. Hansen ML, Gunn PW, Kaelber DC. Underdiagnosis of hypertension in children and adolescents. JAMA 2007;298:874-9.
21. Ealovega MW, Tabaei BP, Brandle M, Burke R, Herman WH. Opportunistic screening for diabetes in routine clinical practice. Diabetes care 2004;27:9-12.
22. Singh H, Thomas EJ, Sittig DF, et al. Notification of abnormal lab test results in an electronic medical record: do any safety concerns remain? The American journal of medicine 2010;123:238-44.
23. 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.
24. Callen J, Georgiou A, Li J, Westbrook JI. The safety implications of missed test results for hospitalised patients: a systematic review. BMJ quality & safety 2011;20:194-9.
25. Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. Annals of internal medicine 2006;145:488-96.
26. Singh H, Hirani K, Kadiyala H, et al. Characteristics and predictors of missed opportunities in lung cancer diagnosis: an electronic health record-based study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2010;28:3307-15.
27. Singh H, Daci K, Petersen LA, et al. Missed opportunities to initiate endoscopic evaluation for colorectal cancer diagnosis. The American journal of gastroenterology 2009;104:2543-54.
28. Schiff GD, Hasan O, Kim S, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. Archives of internal medicine 2009;169:1881-7.
29. Hinton CF, Neuspiel DR, Gubernick RS, et al. Improving newborn screening follow-up in pediatric practices: quality improvement innovation network. Pediatrics 2012;130:e669-75.
30. Miller MR, Griswold M, Harris JM, 2nd, et al. Decreasing PICU catheter-associated bloodstream infections: NACHRI's quality transformation efforts. Pediatrics 2010;125:206-13.
31. Miller MR, Niedner MF, Huskins WC, et al. Reducing PICU central line-associated bloodstream infections: 3-year results. Pediatrics 2011;128:e1077-83.
32. 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;91:354-94.
33. Greevy R, Lu B, Silber JH, Rosenbaum P. Optimal multivariate matching before randomization. Biostatistics 2004;5:263-75.
34. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Grol RP. Evidence for the impact of quality improvement collaboratives: systematic review. Bmj 2008;336:1491-4.
35. Hulscher ME, Schouten LM, Grol RP, Buchan H. Determinants of success of quality improvement collaboratives: what does the literature show? BMJ quality & safety 2013;22:19-31.
36. Gaur AH, Bundy DG, Gao C, et al. Surveillance of hospital-acquired central line-associated bloodstream infections in pediatric hematology-oncology patients: lessons learned, challenges ahead. Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America 2013;34:316-20.

37. Sittig DF, Teich JM, Osheroff JA, Singh H. Improving clinical quality indicators through electronic health records: it takes more than just a reminder. Pediatrics 2009;124:375-7.
38. Singh H, Arora HS, Vij MS, Rao R, Khan MM, Petersen LA. Communication outcomes of critical imaging results in a computerized notification system. Journal of the American Medical Informatics Association : JAMIA 2007;14:459-66.
39. Kaplan HC, Provost LP, Froehle CM, Margolis PA. The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. BMJ quality & safety 2012;21:13-20.
40. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. The American journal of medicine 2008;121:S2-23.
41. Giardina TD, King BJ, Ignaczak AP, et al. Root cause analysis reports help identify common factors in delayed diagnosis and treatment of outpatients. Health affairs 2013;32:1368-75.
42. 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.
43. Rinke ML, Bundy DG, Chen AR, et al. Central line maintenance bundles and CLABSIs in ambulatory oncology patients. Pediatrics 2013;132:e1403-12.
44. Zuckerbrot RA, Jensen PS. Improving recognition of adolescent depression in primary care. Archives of pediatrics & adolescent medicine 2006;160:694-704.
45. Schubiner H, Tzelepis A, Wright K, Podany E. The clinical utility of the Safe Times Questionnaire. The Journal of adolescent health : official publication of the Society for Adolescent Medicine 1994;15:374-82.
46. Force USPST. Screening and treatment for major depressive disorder in children and adolescents: US Preventive Services Task Force Recommendation Statement. Pediatrics 2009;123:1223-8.
47. Zuckerbrot RA, Maxon L, Pagar D, Davies M, Fisher PW, Shaffer D. Adolescent depression screening in primary care: feasibility and acceptability. Pediatrics 2007;119:101-8.
48. Williams SB, O'Connor EA, Eder M, Whitlock EP. Screening for child and adolescent depression in primary care settings: a systematic evidence review for the US Preventive Services Task Force. Pediatrics 2009;123:e716-35.
49. Asarnow JR, Jaycox LH, Duan N, et al. Effectiveness of a quality improvement intervention for adolescent depression in primary care clinics: a randomized controlled trial. JAMA 2005;293:311-9.
50. Tschudy MM, Arcara KM. The Harriet Lane Handbook. 19th ed. United States: Elsevier; 2012.
51. Allgaier AK, Pietsch K, Fruhe B, Sigl-Glockner J, Schulte-Korne G. Screening for depression in adolescents: validity of the patient health questionnaire in pediatric care. Depression and anxiety 2012;29:906-13.
52. Zuckerbrot RA, Cheung AH, Jensen PS, Stein RE, Laraque D, Group G-PS. Guidelines for Adolescent Depression in Primary Care (GLAD-PC): I. Identification, assessment, and initial management. Pediatrics 2007;120:e1299-312.
53. Stein RE, Zitner LE, Jensen PS. Interventions for adolescent depression in primary care. Pediatrics 2006;118:669-82.
54. 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;44:387-8.
55. Expert Panel on Integrated Guidelines for Cardiovascular H, Risk Reduction in C, Adolescents, National Heart L, Blood I. Expert panel on integrated guidelines for cardiovascular health and risk reduction in children and adolescents: summary report. Pediatrics 2011;128 Suppl 5:S213-56.
56. Duncan PM, Duncan ED, Swanson J. Bright Futures: the screening table recommendations. Pediatric annals 2008;37:152-8.

57. Parker E, O'Connor P, Adams K. C-B5-04: Patterns of Blood Pressure Measurement at Pediatric Primary Care Visits in Large Medical Group Practices. Clinical Medicine and Research 2011;9:153-4.
58. Shapiro DJ, Hersh AL, Cabana MD, Sutherland SM, Patel AI. Hypertension screening during ambulatory pediatric visits in the United States, 2000-2009. Pediatrics 2012;130:604-10.
59. 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; 2013 August