Improving the Quality of Pediatric Emergency Care Using an Electronic Medical Record Registry and Clinician Feedback (PECARN Emergency Care Registry) PECARN Protocol Number 030

Pediatric Emergency Care Applied Research Network
Maternal and Child Health, Emergency Medical Services for Children (EMSC) Program

Protocol Version 1.02
Version Date: December 20, 2012
Printing Date: December 20, 2012

This protocol is PECARN Protocol Number 030, and the lead PECARN investigator for this protocol is Elizabeth R. Alpern, MD, MSCE, The Children's Hospital of Philadelphia, University of Pennsylvania.

The PECARN Research Node Centers are the University of Michigan at Ann Arbor, Cincinnati Children's Hospital Medical Center, Columbia University, University of California at Davis Medical Center, Children's Hospital of Pittsburgh, and Children's National Medical Center are supported by Cooperative Agreements U03-MC00003, U03-MC22684, U03-MC00007, U03-MC00001, U03MC22685, and U03-MC00006 from the Emergency Medical Services for Children (EMSC) Program, Maternal and Child Health Bureau, Health Resources and Services Administration. This project is directly supported by R01 HS020270 from the Agency for Healthcare Research and Quality.

This document was prepared by the PECARN Data Coordinating Center (DCC) located at the University of Utah School of Medicine, Salt Lake City, Utah. The document was written and typeset using LaTeX 2_ε. The DCC at the University of Utah is supported by Cooperative Agreement U03-MC00008 from the Emergency Medical Services for Children (EMSC) Program, Maternal and Child Health Bureau, Health Resources and Services Administration.

PROTOCOL TITLE:

Improving the Quality of Pediatric Emergency Care Using an Electronic Medical Record Registry and Clinician Feedback

Short Title: PECARN Emergency Care Registry
PECARN Protocol Number: 030

Lead Investigator and Author:
Elizabeth R. Alpern, MD, MSCE
The Children's Hospital of Philadelphia, University of Pennsylvania

Protocol Version: 1.02
Version Date: December 20, 2012

I confirm that I have read this protocol, I understand it, and I will conduct the study according to the protocol. I will also work consistently with the ethical principles that have their origin in the Declaration of Helsinki and will adhere to the Ethical and Regulatory Considerations as stated. I confirm that if I or any of my staff are members of the Institutional Review Board, we will abstain from voting on this protocol, its future renewals, and its future amendments.

Principal Investigator Name: _____

Principal Investigator Signature: _____

Date: _____

THIS PAGE IS INTENTIONALLY BLANK.

Contents

List of Tables

List of Figures

1 Study Summary

Approximately 20% of all children in the US will have at least one Emergency Department (ED) visit each year. Patients and caregivers expect clinicians to deliver high quality emergency care, yet many children do not receive appropriate care under the current system. The Institute of Medicine report, "Emergency Care for Children: Growing Pains," notes that the delivery of care should be built on a strong foundation in which emergency care is based on scientific evidence, data are collected so clinicians can learn from past experience, and system performance is monitored to ensure quality. While EDs routinely collect substantial patient data with the potential to provide information on the quality of care provided, the capacity to capture, analyze and report these data back to front-line clinicians capable of acting to improve health care delivery has been limited.

The Pediatric Emergency Care Applied Research Network (PECARN) began collecting data for the PECARN Core Data Project (PCDP) in 2002, and this administrative registry now has over 6 million visit records in it. However, the PCDP has not included clinically important data about processes and outcomes of care, changes in clinical status, physiologic measures of severity of illness, laboratory data, or narrative reports. These types of clinical data, while important for improving the care of children in our emergency departments, have been prohibitively expensive to extract from the (largely paper based) medical record on a large scale. Over the past decade, however, available electronic health records (EHR) have improved drastically, and several PECARN hospitals have implemented nearly 100% electronic records. As the state of the EHR advances in the PECARN sites, a unique opportunity exists to build upon this new digital infrastructure, and develop a clinically rich database of electronically abstracted data from the EHR. This database will be called the PECARN Emergency Care Registry.

In this project, clinically meaningful performance measures will be derived from registry data, and variability of practice will be measured among participating sites and the individual clinicians in those sites. These performance measures will be obtained from specific database fields (for example, the length of stay can be calculated from the times of admission and discharge) in many instances. However, innovative natural language processing (NLP) techniques will be used to obtain valuable information from free text narrative fields, which have not been previously used in PECARN studies. For example, the components of the Pediatric Asthma Severity Score or the

Glasgow Coma Scale are often written in the middle of a nursing or physician assessment note, rather than being separated out into specific database fields. By using the Registry with its comprehensive clinical information, it will be possible to derive sophisticated performance measures such as determining how much time elapses between the presentation of a child with a long bone fracture and the provision of pain relief.

Selected performance measures will be combined into useable report cards for each participating site and the individual clinicians in those sites. These Quality Performance Measure Report Cards will show the performance of the site or clinician, but will also include the best scores across all sites and clinical providers. These best scores will be called the benchmark targets, and will enable sites and individual clinicians to try to improve to the very best possible performance. The hypothesis of the study is that providing this type of information to individual clinicians (who are the only people who can truly affect meaningful changes in care) will result in improved performance and decreased variability (variation) of care for an individual clinician, between different clinicians, and eventually between hospital sites.

The project has enormous potential to improve our ability to evaluate systems of health care delivery, as well as to lead to improvements in the quality of care provided to acutely ill or injured children. Although this protocol centers on health care delivered to children in PECARN emergency departments, the principles and lessons that we learn will extend to any component of the healthcare system that collects electronic patient data. This project will have wide applicability for clinical quality improvement, comparative effectiveness research, and indeed, clinical research, across all healthcare domains.

1.1 Study Objectives

This study has three objectives:

Objective 1. Develop an emergency care visit registry for pediatric patients by merging electronic health record (EHR) clinical data from participating hospital emergency departments (ED) for Quality Improvement purposes and to support future research.

Objective 2. Use the emergency care visit registry to collect stakeholder-prioritized emergency care performance improvement measures for important pediatric medical and trauma conditions at the level of the ED and individual clinician.

Objective 3. Report emergency care performance improvement measures to individual ED clinicians and to sites, and measure subsequent changes in quality performance.

2 Background

Children account for over one quarter of the 114 million annual emergency department (ED) visits in the US,[1, 2] yet the needs of children in the ED have received relatively small attention.[3] Recently, there has been increased attention to the unique needs of children during emergencies. Among the areas in need of evaluation are the effectiveness and quality of emergency care, outcomes of different configurations of EMSC, optimal resource allocation and utilization, and cost-effectiveness of EMSC and its components.[4] There has also been documentation of significant variation in the quality of care provided to children in EDs,[5–11] as well as recognition of racial and ethnic disparities in healthcare delivery.[12] The Pediatric Emergency Care Applied Research Network (PECARN) was established in 2001 to address these needs and improve the quality of emergency care for children through the performance of rigorous research.[13, 14]

The network's first project was to design and implement an administrative database of all emergency department visits in PECARN hospitals, the PECARN Core Data Project (PCDP).[15] The PCDP, which has been ongoing with Institutional Review Board (IRB) approval at every PECARN hospital since 2002, uses data from billing and registration systems at all participating hospitals, as well as patient tracking data from some hospitals. The PCDP currently has over 6 million visits to the PECARN sites. The project has been successful in providing basic demographic and injury or illness information,[15–18] evaluation of methodology,[19, 20] developing diagnosis grouping and severity systems,[21, 22] and providing limited benchmarking information.[16–18, 23]

In the first year of the PCDP, investigators also collected detailed clinical information for a sample of children (10,000 total) seen in PECARN emergency departments by linked manual chart review. Availability of clinical data, in addition to the administrative data available in the overall PCDP database, allowed investigators to evaluate practice pattern variation for asthma and long-bone fractures.[10, 24] Pain was under-documented in children with long bone fractures and only two-thirds of these patients received analgesics in

the ED.[24] More than one-third of patients treated for status asthmaticus received potentially unnecessary ancillary testing and the risk for this testing was higher in children cared for at non-children's hospitals and by clinicians without subspecialty training in pediatric emergency medicine.[10] The current project builds off of this early experience, but leverages the availability of electronic health records (EHR) in selected PECARN sites. The PECARN Emergency Care Registry, which will be constructed in this project, will have detailed clinical data about all children seen in the ED, allowing performance improvement for the entire spectrum of problems encountered in this setting.

3 Preliminary Data

This project leverages previous experience with the PCDP, but extends the application of sophisticated biomedical informatics to the pediatric emergency setting. The major informatics innovation of this project is natural language processing (NLP) of free text fields such as nursing and physician assessment notes, free-form physical examination findings, etc. Indeed, the majority of useful clinical information probably is contained in these free text (narrative) notes. NLP will be used to identify key clinical information within these notes, such as the components of a pain scale score or asthma severity score.

A pilot study was done to demonstrate the feasibility of using NLP to automatically obtain useful information from narrative notes. To determine whether or not anti-seizure medications were administered in a timely fashion, it is necessary to identify when seizure activity was first noted. Similarly, for asthma severity scores, it is possible that key data elements such as severity of wheezing may be documented only in narrative portions of the physical exam. In this pilot study, investigators obtained a sample of de-identified charts for these two diagnoses (seizures and asthma) from three of the PECARN hospitals that are participating in this project.

3.1 Identifying Status Epilepticus

Sixty-seven records of children with status epilepticus were used for analysis. This sample was reviewed manually by the investigators to identify the cases where seizure activity was present at the time of arrival in the ED. Several automated NLP algorithms (methods) were then used to classify the presence or absence of seizure activity on arrival. Each algorithm used a combination of triage acuity, presence of a specific diagnosis code for status epilepticus, and phrases specifically indicating the presence or absence of

seizure activity. Negation was handled by looking for the words "no","not", "without", or "absent" within four words before a term of interest. The presence or absence of seizure activity on arrival was correctly classified for 62/67 (93%). In contrast, using only administrative data (triage acuity and discharge diagnosis) and no NLP correctly classified only 20/27 (74%). These data show that NLP algorithms can find relevant information in the narrative text, which will enable us to create more meaningful performance measures than if much of the information was missing or incorrectly classified.

3.2 Identifying Asthma Severity Scores

Fifty-nine records of children with asthma-related diagnoses were used for this analysis. At one institution (19 records), the asthma severity scores were already systematically recorded in structured database fields, facilitating direct comparison with NLP results. In 18 records, a second asthma severity score could be constructed with NLP techniques. In 40 records from the other two pilot institutions, NLP algorithms looking for key phrases about the lung examination found adequate descriptions of wheezing, respiratory effort, and expiratory phase to calculate a pre-treatment asthma severity score[25] in 35 (88%). These results provide another example of how NLP can enhance our ability to extract useful information from narrative notes with a computer, and create meaningful performance measures with which to provide useful feedback to clinicians..

4 Overview of Study Design

Investigators will build the PECARN Emergency Care Registry (henceforth referred to as the Registry) using EHR data for pediatric ED visits at each site. Electronic data from each site will be transmitted to the PECARN Data Coordinating Center (DCC) at the University of Utah.

Performance measure outcomes (Table 1 on page 16) will be derived from the Registry data and descriptive statistical analyses will include appropriate parameter estimates and measures of variability among sites and individual clinicians. "Achievable benchmarks of care" (ABC^{TM}) will be determined using established methods.[26–28] Investigators will use both the ABC^{TM} as well as expert panel determination of the ideal benchmarks of care. An expert panel will review the achievable benchmarks and determine ideal benchmarks for each performance improvement measure. The expert panels will be informed by the quantitative baseline variability of measures that

have been derived from the Registry, as well as by published literature and their own expert opinions.

From this expert panel meeting, investigators will design site level and clinician level Quality Performance Improvement Measures Report Cards that will be used for Objective Three. Site level report cards will contain all 14 performance improvement measures with selected descriptive statistics (e.g., means, medians, standard deviations or proportions with 95% confidence intervals for rates and proportions), and benchmarks noted. Clinician level report cards will contain information from a selection of the 14 performance measures including but not limited to the seven specific performance improvement measures (shaded in Table 1 on page 16) that are within the locus of control of the individual clinician and, thus, amenable to physician behavior change.

During the subsequent 24 months, data will be extracted for every pediatric patient visit to the ED in each of the participating sites, added to the Registry, and the Quality Performance Improvement Measures Report Cards will be produced. These Report Cards will be distributed to each site and clinician. It is anticipated that there will be a delay of up to three months between clinical events and Report Card production and distribution.

The seven specific measures that are provided to clinicians, which address common and important clinical diagnoses in pediatric emergency medicine, will be evaluated for improvement in the measures themselves as well as decreased variation in performance within and between sites and groups of clinicians. After three months of Report Cards have been distributed, investigators will discuss and solicit feedback from sites and clinicians to assess the acceptability of format and information contained within the Quality Performance Measures Report Cards. Modifications to the Report Cards will be guided by this information.

5 Study Population

This study will assess performance measures for each of the participating hospital sites. The goals are to reduce differences between high performing sites and those with opportunities for improvement, and to achieve benchmark performance levels at all participating sites. The four sites (plus their satellite hospitals, for a total of eight facilities) have approximately 300

clinicians whose performance will contribute to overall site performance. Thus, the subjects that are under study in this project are primarily sites and secondarily their clinicians.

To have meaningful performance measures for sites and clinicians, there must be sufficient numbers of patients seen with the respective conditions of interest. Based on calendar year (CY) 2010 information from the participating sites, there will be ≈416,000 annual visits for patient populations. The total patient visits during the study will be ≈1.3 million (CY 2012 plus 24 months of feedback data). These visits will include ≈93,000 asthma visits, ≈30,000 long bone fracture visits,≈1,700 diabetic ketoacidosis (DKA) visits, ≈1,900 status epilepticus visits, and ≈183,000 visits for viral infections. These numbers are all estimations; the database will include all patient visits for CY 2012 (used in Objectives One and Two) in addition to all visits occurring during the 24 month feedback period for Objective Three.

6 Study Procedures

6.1 Database Construction

6.1.1 Summary of Process

There are three EHR vendors in use at the PECARN sites that will participate in this project. Implementation of the Registry will occur in the following general stages:

1. Identify potential sources of relevant data elements in the specific EHR at each site.
2. Finalize the types of data elements that will be extracted.
3. Extract data for one day of data at each clinical site.
4. Transmit one day data to the DCC for de-identification.
5. Establish de-identification procedure at each clinical site.
6. Extract and de-identify one month data from calendar year (CY) 2012 at each site.
7. Transmit one month de-identified data to DCC from each site.

8. Finalize and test import procedures from one month extracts into Registry.

9. Analyze frequencies of missing, out of range, or unexpected values for key data elements.

10. Extract, de-identify, and transmit entire CY 2012 from each site to the DCC.

11. Create Registry with entire CY 2012 from all participating sites.

6.1.2 Data Domains of Interest

The EHR contains a large number of data, including information required for auditing the medical record, and it is not intended to extract all EHR information into the Registry. For example, if a laboratory value is incorrectly entered, and then corrected, both values are retained in the EHR for legal auditing purposes. The Registry should only receive the final value. The following data domains are expected to be extracted from the EHR:

Demographic Data. For example, but not limited to: site of care, insurance type, race, ethnicity, birthdate (age).

Clinician Data. For example, but not limited to: independent licensed provider.

Date and Time Data. For example, but not limited to: all date and time values relating to the ED or hospital visit, as well as date and times of all events or findings.

Diagnoses and Procedures. For example, but not limited to: all available diagnoses codes, procedure codes, cause of injury codes. Potentially ICD9, ICD10, and CPT.

Review of Systems. For example, but not limited to: review of physical systems as recorded in specific fields, if available. Otherwise derived from narrative text data.

History of Illness. For example, but not limited to: history of illness as recorded in specific fields, if available. Otherwise derived from narrative text data.

Physical Examination. For example, but not limited to: physical examination findings (including weight) as recorded in specific fields, if available. Otherwise derived from narrative text data.

Laboratory Testing Data. For example, but not limited to: all laboratory tests sent during the ED visit, including results that may be returned after the ED visit is concluded.

Medications Ordered and Administered. All medications ordered or administered in the ED, including dosage information.

Vital Sign Data. All vital signs as recorded in specific fields, if available. Otherwise derived from narrative text data.

Narrative Text Data. All available narrative data related to the ED visit, including potentially delayed radiologic dictation reports. Narrative or free text notes may be created by physicians, nurses, or other clinicians, and the author of each narrative will also be extracted.

ED Admission and Discharge Data. For example, but not limited to: Reason for visit, chief complaint, triage category, mode of arrival, urgency, discharge location, vital status.

Inpatient Admission and Discharge Data. For ED visits resulting in a hospital admission at the same hospital, the date and time of hospital discharge, vital status at discharge, and hospital discharge summary.

Data within these domains of interest will then be used to derive the performance improvement measures to be used in this project. These are shown in Table 1 on the following page.

For purposes of this research project, data that are not available within one month of the ED visit will be considered unavailable or missing, and no effort will be made to follow up for these delayed data. For example, if a patient has a radiologic procedure but the dictation is not in the EHR within 30 days of the visit, that dictation will not be incorporated into the Registry. This time window is necessary because of the logistical issues of attempting to follow up every potential result, as well as the time constraints relating to producing timely feedback reports for Objective Three.

Table 1: Performance measures (clinician measures shaded)

Performance measure	Numerator	Denominator
Measuring weight in kilograms	Number of emergency department visits with a weight in kilograms documented during the current ED visit	Number of emergency department visits
Measuring vital signs for ED patients	Number of visits with all 4 vital signs documented (temperature, heart rate, respiratory rate and blood pressure) by the completion of the first nursing assessment (e.g. triage or room placement)	Number of emergency department visits
ED Door to Clinician Time	Time interval between patient presentation to the ED and the first time the patient is seen by a clinician, excluding triage personnel, who can initiate a diagnostic evaluation or therapeutic plan for all patients	Include all emergency department visits; Exclusion: Left Without Being Seen patients
Total ED Length of Stay	Time from ED arrival to ED departure	Include all emergency department visits; Exclusions: Left Without Being Seen, Left Without Treatment and Left Against Medical Advice
ED Left Without Being Seen Rate	Number of visits where a patient left without being seen by a clinician who can initiate a diagnostic and therapeutic plan, excluding triage personnel	Number of emergency department visits
ED Return Visits within 48 hours resulting in admission (rate)	Number of patients returning to the ED within 48 hours of a prior ED visit whose return visit results in hospital admission	Total number of ED discharges
Plain film imaging turnaround time: time to image available to ED staff	Time interval between plain film order and image available for viewing by ED staff	N/A
Reducing pain in children with acute fractures	Number of patients with pain assessed and reassessed using the same age-appropriate pain scale who show documented improvement in pain score within 90 minutes of ED arrival	Number of patients with acute long-bone fractures

continued on next page

Table 1: Performance measures (clinician measures shaded) (*continued*)

Performance measure	Numerator	Denominator
Timeliness of insulin administration for patients in diabetic ketoacidosis	Time from ED arrival to insulin administration for patients in diabetic ketoacidosis	N/A
Timeliness of treatment with anti-epileptic drugs for patients in status epilepticus	Number of patients receiving an anti-epileptic drug within 10 minutes of ED arrival	Number of patients presenting to the ED in status epilepticus
Systemic corticosteroids in asthma patients with acute exacerbation	Number of asthma patients receiving systemic corticosteroid during ED visit	Number of patients with a diagnosis of asthma treated with >1 inhaled β-agonist, age $\geq$2 years
Timeliness of inhaled β-agonist treatment for patients with acute asthma exacerbations	Time from ED arrival to first inhaled β-agonist administered in the ED	N/A
Objective improvement in asthma severity score for patients with acute asthma exacerbations	Number of patients whose objective Pediatric Asthma Severity Score (PASS) at discharge is less than the PASS at presentation	Number of patients with a diagnosis of asthma treated with >1 inhaled β-agonist, age $\geq$2 years
Reducing antibiotic use in children with viral illnesses	Number of eligible patients given antibiotics in ED or discharged with an antibiotic prescription	Number of patients with an discharge diagnosis of URI, viral illness, viral syndrome, or fever and who do not also carry a diagnosis of bacterial infection

6.1.3 Location of Data within EHR

For each site, the principal investigator of this proposal (Dr. Alpern), site principal investigator, site health information technologist, and DCC staff will undertake a review to identify and verify the locations within each EHR for each data domain of interest for an ED visit at each site. A list of specific EHR locations for each data domain of interest will be compiled and used as the source of the data element at each site, with particular attention to the variables required to compile the performance measures shown in Table 1 on the preceding page.

To construct performance improvement measures, a single defined source for each variable will be identified, if possible. When multiple sources may exist for a single variable, investigators will define a source hierarchy (e.g. look in the attending physician note first, then check the nurse assessment note,

then check the triage note). For example, multiple observers may record an asthma severity score, and it may exist either as a free text description, or as a structured variable in the EHR. For these data to be comparable across sites, it may also be important to specify whether the information was recorded by a physician, nurse, or respiratory technician.

6.1.4 Database Extraction

Available data will be extracted from each site's EHR in an automated fashion that is anticipated to be vendor-specific, or even installation specific (i.e. two different sites may have the same vendor but different preferred extraction methods depending on the vendor features that were purchased or updated). It is anticipated that the final extraction form will be plain text formatted with extensive markup language (XML) tags. The XML format will facilitate validating the extraction data, and by using plain text, de-identification procedures (Section 6.2) can readily be applied to the extracted information prior to importing data into the Registry database.

The precise manner of database extraction, formatting of variables, and technical details of the steps required to accomplish this task are described here for information purposes. The technical approach may be altered without revising this protocol document.

6.2 Deidentification Procedures

The text file that is produced by automatic extraction will be fully identified, and sophisticated algorithms will then be used to remove essentially all HIPAA-defined identifying information. There are only three potentially identifying data elements that will remain in the study record after the de-identification procedure is carried out: the site identification, the clinician identification, and the month of the visit. These three data elements are needed in order to provide feedback for specific months to specific sites and clinicians. All other patient-related identifiers will be removed by the de-identification procedure.

The precise technology to be employed may be changed as this project moves forward to reflect improvements in available tools, without revision of this protocol. The purpose of this section is not to constrain the use of state-of-the-art technologies, but rather to describe the principles by which the process will be carried out. The DCC will work with project informaticists

and provide tools for each clinical site to employ prior to transmitting the full CY 2012 data. In order to finalize and verify the effectiveness of the de-identification, the DCC will have access to fully identified records from each site from a single day of 2012 used for site specific De-ID algorithm derivation. After the de-identification procedures have been confirmed, all these identifiable records that are at the DCC will be irrevocably destroyed and will not be incorporated into the Registry. Subsequent de-identification will be carried out by each hospital prior to transmission to the DCC.

The de-identification is planned to be done with software (De-Id) that has been developed by other investigators[29] to produce a completely de-identified public use intensive care database.[30] There are other potentials tools for the same purpose,[31–33] and it may become preferable to change the software as these tools evolve.[34]

6.2.1 General Algorithm

The software (De-Id) is a program script that uses multiple dictionaries to identify and remove names and locations from text. A list of known patient names (at an individual hospital) can be the first dictionary used at the site by the script, and this will remove 100% of patient names. The software then uses large dictionaries of first and last names to identify other potential names, such as relatives, nurses, doctors, or other clinicians. A similar approach is used for locations — namely, a list of locations that are relevant to each clinical site will be used as a location dictionary, and locations will be removed.

Dates and times present an interesting problem because it is necessary to preserve time intervals and patient ages. The software will remove all dates and replace them with a random patient-specific offset. This shifted date preserves the day of the week and the season to prevent confusion due to free text phrases such as “next Wednesday”. The patient’s date of birth is treated in the same manner, so that the correct age can be calculated from the replaced dates. The random shift is patient specific and will be consistent for the entire individual patient record and between subsequent visits for that patient.

In order to preserve the month in which the data were submitted, the clinician identification,and site identification, separate database fields will be added to each record when the file is sent to the DCC, so that month-specific feedback

can be provided to sites and providers, as already described.

Other identifiers include telephone and fax numbers, Social Security numbers, medical record numbers or other numeric identifiers. The software will remove these throughout. The medical record number will be one-way hash encrypted to allow identification of repeat visits for patients.

6.2.2 Effectiveness of Algorithm

This software has been tested rigorously.[29] Three human de-identifiers manually de-identified text for comparison with the software. On test data consisting of nearly 2,000 free text nursing notes, the software performed better than the average individual human de-identifier, better than the best single human de-identifier, and better than the average consensus of two human de-identifiers. The software has subsequently been used to de-identify the MIMIC II intensive care database previously mentioned,[30] processing approximately 700,000 nursing notes, 30,000 inpatient discharge summaries, and 300,000 radiology reports, containing a total of over 220 million words. A sample of de-identified text is shown in Figure 1 on the next page.

6.3 Natural Language Processing (NLP) Procedures

Most of the data needed for the performance measures in Table 1 on page 16 are readily available from discretely recorded data elements within the EHR, such as time stamps and computer order entry of medications. However clinicians do not and probably never will enter every data element discretely at the point of care. This project will use natural language processing (NLP) and free-text parsing methods because it is clear that important knowledge about ED patients remains in free text. NLP will provide access to the richness of the EHR free text to overcome significant limitations of discrete variables; for example, lack of depth to the ICD-9 diagnosis code fails to capture the state of the patient at clinical presentation in status epilepticus. NLP will allow the capture of all of the elements of patient-centric measures such as reduction in pain and improvement in respiratory status. NLP has not been widely used by ED researchers, but its increasing application in other clinical disciplines supports its relevance here.[35–38]

In several identified instances (pain scores, asthma severity scores, and certain diagnoses), the investigators will benefit from using free text data rather than discrete data fields, either because the current EHR interface

==========

DISCHARGE SUMMARY

Name: [**Known patient lastname**], [**Known patient firstname**]

[**Unit Number 626**]

Admission Date: [**2016-11-07**]

Discharge Date: [**2016-11-22**]

Date of Birth: [**1972-09-20**]

Sex: F

HISTORY OF PRESENT ILLNESS: Patient is a 44-year-old lady status post living related kidney transplant on [**2016-10-19**], who presented at [**Hospital 36**] for end-stage renal disease secondary to type 1 diabetes mellitus.

She presented to [**Hospital1 **] on [**2016-11-07**] with increased drainage from her surgical wound and JP, increased abdominal pain, and anuria x4 days. The patient reported constipation for a week. She denies flatus. She was complaining of nausea and vomiting. Her abdominal pain had become progressively worse left lower quadrant most notable. There is no radiation to the back or elsewhere. She denied any fevers, chills. She noted decreased p.o. intake recently. Her drainage from her wound incision and JP was notable for yellowish clear urine smelling fluid.

==========

Figure 1: Example of de-identified narrative text[29]

does not support directed data fields or because some clinicians will chart using free text. Bioinformatics investigators will use NLP techniques and develop algorithms to extract text fragments that are relevant to each specific measure, and normalize the extracted text to a controlled vocabulary for each measure. Controlled (limited) vocabularies of terms are commonly used to indicate particular attributes (e.g. presence or absence of seizures). In these situations a pattern-matching approach using regular expressions and simple tactics to identify potential negation are often sufficient to accurately classify variables from narrative documentation.

Based upon the pilot assessment of free text fields (Section 3 on page 10), it is likely that relatively simple NLP techniques, such as simple pattern matching, will be required. Much more complicated methods can be used, if necessary. These include having investigators manually annotate (code) information in some charts, and then training a computer to correctly identify the coded information. In technical terms, this is called machine learning, and involves building a probability-based extractor of key phrases. This means that the NLP program uses probabilities to identify key phrases. Some of the NLP methods that can be used include conditional random fields, maximum entropy, and support vector machines.

To normalize extracted text, a custom rule-based program will align the extracted phrases with a vocabulary list for each measure. This could be performed by using a prior available or expert-constructed (most likely) controlled vocabulary for each measure. Extracted text will be closely scrutinized for errors by domain experts for each measure, and an iterative development process will optimize accuracy.

6.4 Determining Benchmarks for Report Card

Investigators will use both the "Achievable Benchmark of Care" (ABC^{TM}) methods[26–28] as well as expert panel determination of ideal benchmarks of care. This will allow us to identify situations where the achievable benchmark of care still demonstrates a significant gap in respect to ideal performance. We will convene an expert panel that includes national leaders in general and pediatric emergency medicine, quality improvement, patient safety, and performance measurement. Consensus methods provide a means of synthesizing information where unanimity of opinion does not exist owing to insufficient, or conversely an overload, of information. In order to overcome these issues in the development of the performance measure benchmarks,

investigators will use Nominal Group Techniques. These methods attempt to assess the extent of agreement (consensus measurement) and to resolve disagreement (consensus development). Strengths of these methods include 1) *Anonymity*: dominance is avoided by using private ratings in Nominal Group; 2) *Iteration*: processes occur in "rounds", allowing individuals to change their opinions; 3) *Controlled Feedback*: the distributions of the group's responses are shown; and 4) *Statistical Group Response*: judgment is expressed using summary measures of the full group response, giving more information than just a consensus statement.

Prior to the expert meeting, working group members will be sent information including a project overview and data derived from the Registry containing data from CY 2012. Each performance measure will have data derived from the Registry and descriptive statistics (percents, means, medians, standard deviations, etc.) for the measures will be provided. The ABC^{TM} identifies the "best practice by a few superior clinicians"[28] by using the "pared mean." The pared mean is the average performance of a subset of clinicians with the highest scores for the measure. This subset is comprised of the top-ranked clinicians at the point that 10% of the patient pool, across all the clinicians, is included. This will be determined for all clinicians together as well as for each site using only clinicians at that site. The ABC^{TM} will be used as the starting framework for discussion by the Expert Panel.

The Expert Panel meetings will utilize the Nominal Group Techniques to determine if there should be any difference between the ABC^{TM} and the "ideal benchmark of care." Prior to this first meeting, all participants will be asked to record, individually and without discussion, their own ideal benchmark for each of the quality performance measures informed by the ABC^{TM} generated from Objective One. At the meeting, each individual, in a round-robin fashion, will present their ideal benchmark for each performance measure. In the next phase, a discussion of the benchmarks will occur. Participants will evaluate each performance measure separately and, when necessary, clarify their ideas. After the discussion, each participant, privately and in writing, will rate the various proposed ideal benchmarks for each performance measure. The group's views will be tabulated by the investigators during the session. Each performance measure will be given several minutes of discussion, after which expert panel members must indicate their acceptance of the ideal benchmark for each performance measure. Disagreement will be resolved at this time.

6.5 Report Card Feedback

Having determined the ABC^{TM} from the Registry and set the ideal benchmark by expert consensus, the Quality Performance Measure Report Cards will be generated from the Registry on a monthly basis. Report Cards will be generated on a site level and will also be generated on a clinician level. The site-specific Report Card will contain the site mean, median, and standard deviation for each performance measure. The ABC^{TM} as well as the ideal benchmark for each measure will also be provided for comparison. In addition, each site will receive feedback using statistical process control charts. The type of control chart for each measure will be determined using standard methods dependent upon the type of data being plotted (e.g. continuous data such as ED length-of-stay versus categorical data such as whether an asthma patient received a corticosteroid during their ED visit).[39] The ABC^{TM} as well as the ideal benchmark for each measure will also be provided for comparison on each control chart. In addition, the report card will provide the results for the other sites (blinded) for comparison.

The clinician-specific Report Card will contain that practitioner's individual results from a selection of the 14 performance measures including but not limited to the seven performance measures shaded in Table 1 on page 16 with comparison to site's overall results, anonymous other individual clinicians' results, as well as the benchmarks for each measure. The DCC will replace all individual clinician identifiers in the Registry with a randomly generated study identifier. The Registry and the report cards will only contain this study identifier that will be known only to the individual clinician and within the DCC study team. The DCC will maintain a study key to allow for appropriate identification of visits pertaining to each clinician and also to allow for distribution of the report cards. Qualitative assessment of acceptability of the report cards on a clinician and site level will be solicited from the clinicians and Divisional leadership after the first three months of feedback. This qualitative assessment will allow for modifications to the report card to be implemented as indicated.

7 Statistical Analyses

7.1 Hypothesis

The hypothesis of this study is that providing regular performance improvement measure feedback, derived automatically from the emergency care visit

registry, will improve performance quality and decrease variation within and between participating sites and clinicians.

7.2 Objective One

Process Measures. During the construction of the Registry, investigators will track several process measures to assess the effectiveness and efficiency of the formation of the Registry. For example, measures will include (but are not limited to) the following:

- Total time to generate the data upload for each site
- Total time spent by the DCC and investigators examining and cleaning data
- Number of discrete variables and free text (narrative) fields
- Proportion of missing data for each variable
- Proportion of impossible or nonsensical data for each variable
- Proportion of outlier (out of range) data for each variable

Missing Data. Based on prior work using the PECARN Core Data Project, it is anticipated that nearly all variables of interest will have some missing values.[19] While some of the variables will be nearly complete, the cumulative effect of a small number of missing values in many different variables may lead to a substantial number of records with incomplete information. NLP techniques may reduce the amount of missingness in key variables, but this is not known.

While complete-case analysis of data where values are missing completely at random generally leads to unbiased estimates, it is likely that at least some of the variables will not be missing completely at random. Multiple imputation methods[40–42] may be considered, though these methods are not applicable if data are missing in a totally non-random manner. Multiple imputation estimates a distribution of the missing value based on the values of other, known characteristics. A small set (typically five) of imputed datasets are then created, each with plausible values for the missing data elements. Traditional statistical techniques (e.g., linear regression) are performed on each imputed dataset, and the results from these separate analyses are combined. The final results are statistically valid under reasonable assumptions and reflect the uncertainty due to missing values. Imputation is implementable via multiple approaches depending on the pattern of missing data.

Accuracy of Derived Data. Data abstraction from health records is subject to errors of both omission and commission.[20] Once procedures have been determined for extraction of the desired data elements from both discrete and free text fields, investigators will measure the accuracy of these procedures. A random sample of de-identified records will be selected and reviewed independently by two investigators; each data element will be assigned by the investigators based on manual review of the medical record. If there is agreement between reviewers, this assignment will be considered the criterion standard to which the NLP algorithm-based assignment will be compared. If there is disagreement, a third investigator will review and consensus will determine the standard.

NLP is being used for two purposes: identification of subpopulations of patients in which a specific performance metric will be used (e.g., patients with status epilepticus or patients with acute asthma); and measuring object changes in clinical condition by extracting clinical measures from free text narrative fields (such as changes in pain scores or asthma severity scores). The primary outcome of interest is sensitivity and specificity of categorization by the NLP algorithm compared with expert review. Assuming an agreement value of 95%, a sample of 150 charts will allow calculation of sensitivity with a lower 95% confidence limit of 90%, the minimum acceptable value. To assess accuracy of subpopulation identification by NLP, the sensitivity and specificity of agreement with the disease of interest will be measured. For the second purpose, assessment of changes in patient condition, separate sampling is needed. A random sample of 150 charts initially categorized in each of the five target diagnoses (status epilepticus, DKA, acute asthma, long bone fracture, and viral upper respiratory infection) will be selected for expert review. Another sample of 150 charts regardless of initial computer diagnosis assignment will also be selected, to serve as the basis for calculating specificity. Diagnosis-relevant data elements will also be abstracted by investigators (pain scores for children with long bone fractures, or asthma scores from asthma visits.) Investigators will measure both the sensitivity/specificity and the agreement of NLP derived information with the criterion standard. For example, it is important that the NLP is able to detect a change in pain severity after treatment with high sensitivity and specificity, but it is also important that the algorithm correctly identifies the direction of change (i.e. improvement, worsening, or no change).

7.3 Objective Two

Descriptive statistics of the performance measures will be stratified both by site and by individual clinicians within sites. In addition to measures of completeness, such as proportion of patients with missing data, statistics will include means with standard deviation or medians with interquartile range for continuous variables, and proportions with 95% confidence intervals for rates or proportions measures. The ABC^{TM} for each performance measure within sites and for the Registry as a whole will be determined using the pared mean method.[28] The ABC^{TM} uses Bayesian estimators to calculate the adjusted performance fraction (APF) to reduce the effect of some clinicians having small numbers of eligible patients for some performance measures. This allows all clinicians, even those with small number of eligible patients to be used in calculating the ABC^{TM}. The APF is calculated as the (actual number of patients receiving performance measure +1)/(total number of patients for whom performance measure is appropriate + 2). The APF will be used only to rank order clinicians in descending order of performance for each performance measure. We will then determine the subset by beginning with the number of eligible patients for the highest ranked clinician and proceed to add clinicians until we have accounted for 10% of all patients across all clinicians (for the entire Registry as well as by site). We will then use the actual performance measure data from the Registry (not the APF) to determine the ABC^{TM}. The ABC^{TM} is calculated as the (total number of patients in subset receiving the performance measure)/(total number of patients in subset). Results will be reported in graphical format for each site's performance against the other sites in a blinded manner. Differences in the selected performance measures between sites will be reported. Within sites, individual clinicians will be plotted against other clinicians, also in a blinded manner.

7.4 Objective Three

Analysis of Change in Performance Measures. The effects of providing regular feedback on the attainment of performance measure benchmarks will be measured using statistical process control charts,[43] at the site and clinician levels. This type of chart generally includes upper and lower control limits, plotted above and below the center line, and is used to make determinations about changes in the process from which data is being gathered. The baseline for each measure will be obtained by plotting at least 6 to 12 pre-intervention data points. Well established statistical process control rules

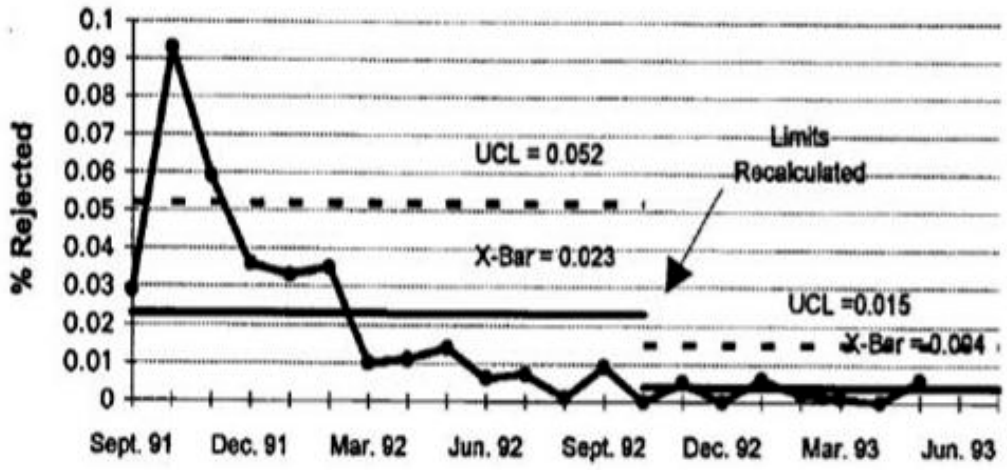

Figure 2: Sample statistic process control chart (industrial)

will be used to detect significant change.[39, 44] An example of a statistical process control chart is shown in Figure 2.

Analysis of Variability Between Clinicians. In order to assess the change in variability in a performance measure across clinicians, the overall rate or average for each clinician prior to and following the intervention will be compared. The pre- and post-intervention variances will be compared using the Ansari-Bradley test. The results of this test must be interpreted with caution, as observed differences may be the result of differences in the distribution of performance measures other than differences in variability. Changes in variance over time (monthly) will also be assessed with particular attention to whether there was a change related to the intervention (feedback).

Generalized Linear Model. In addition to process control charts, which are easily interpretable by clinicians, changes in performance measures (associated with report cards) will be measured with a generalized linear model. This type of model is appropriate due to the longitudinal nature of the data and possible correlation between values within each site and clinician. In some cases, it will be necessary to use a model for non-normal outcomes. This will be determined prior to implementation of the report card. The model

will include a line fit to the data obtained prior to report card implementation and a line fit to the report card period. We will test for both a change in slope and a step-change at the time the report card is implemented. The by-site, by-clinician, and overall fits of this model will be reported. The effects will be tested at the 0.05 level (two-sided). The model will also be constructed in a leave-one-out analysis to determine whether the overall effect is dominated by any particular site or clinician.

Controlling for Secular Trends. One of the difficulties of a pre- versus post-intervention design is the potential for the impact of secular trends independent of the study intervention. Two methods will be used to test for this effect. First, the slope of the change in each performance measure over time will be compared before and after the intervention at each site. A secular trend independent of the intervention will cause a significant non-zero slope. Even in the presence of a non-zero slope prior to the intervention, it is usually possible to detect the effects of the intervention because the slope steepens and the variance (as measured by control limits) is reduced. Second, the feedback intervention will be introduced in a staggered fashion in 2-month intervals all starting 6 months after baseline data collection at each site. For example, Site 1 will start to receive feedback in Month 31, Site 2 in Month 33, Site 3 in Month 35, and Site 4 in Month 37. This will allow Sites 2, 3, and 4 to serve as contemporaneous controls when Site 1 receives the intervention. Similarly, Sites 3 and 4 will serve as contemporaneous controls when Site 2 receives the intervention and Site 4 will serve as a control when Site 3 receives the intervention.

Sample Size. With the currently proposed sites, greater than 1,380,000 ED visits should be available for analysis. This will allow adequate power to detect even small differences across sites and other variables of interest. Specific detectable differences between sites for the primary performance measures are specified in Table 2 on the following page. All power calculations are based on two-sided significance tests with overall $\alpha = 0.05$. A conservative Bonferroni correction was used to adjust for comparisons of each site to a reference site and assumed approximately 10% missing data for assessment of asthma severity score and 5% missing data for all other outcomes.

Smaller differences will be detectable in unadjusted comparisons between sites. Some of these are smaller than differences that would be considered clinically important. This highlights the need to focus on the magnitude of

Table 2: Minimal detectable differences with 80% power

Outcome	# of visits	Estimated proportion or mean	Absolute Difference Detectable	Relative Difference Detectable
Reducing pain in children with acute fractures	430-3300	58-78%	3.7-11.6%	4.7-20.0%
Systemic corticosteroids in asthma patients with acute exacerbation	970-10,000	65-80%	2.0-7.4%	2.6-12.8%
Timeliness of inhaled β-agonist treatment for patients with acute asthma exacerbations	970-10,000	63-70 mins	2.5-8.9 mins	3.5-14.1%
Objective improvement in asthma severity score for patients with acute asthma exacerbations	970-10,000	21-41%	2.1-6.9%	5.9-32.9%
Reducing antibiotic use in children with viral illnesses	2700-18,000	35-43%	1.8-4.7%	4.2-13.1%

observed differences, rather than rely solely on statistical significance. While adjustment for demographic variables may decrease differences between sites, this is also likely to decrease the standard error, allowing for detection of even smaller differences. For the test of change in variability across clinicians, we will observe the main performance measures for over 300 clinicians. Estimates show that there is over 95% power to detect a relative reduction in standard deviation of 25%.

8 Data Management

8.1 Data Transmission to DCC

Data that will be transmitted to the DCC fall into several categories. The DCC uses eRoom^{TM} for certain type of communications; secure file transfer protocol (sFTP) is most often used for large file transfer; discrete study related data are entered into a web based electronic data capture system (currently OpenClinica). All these systems incorporate 128 bit encrypted transmission as described in Section 8.4 on page 32. The most efficient means of secure transmission may change between early data transmission and the monthly transmissions occurring for the last 24 months of the study, but in

no case will data be transmitted without at least 128 bit encryption. No study data will be transferred via email.

8.2 Database at DCC

The DCC will use Microsoft SQL Server and/or PostgreSQL databases to support research activities of this project. These databases provide sophisticated role based security facilities. The databases are not directly accessible from outside the offices of the DCC unless a VPN connection is established. Non-DCC investigators in this project will access the Registry data via a VPN, and will have read-only access to the database.

8.3 Data Storage and Backup

Currently the DCC manages over 1.5 terabytes of storage. with a consolidated storage utility using Dell storage systems and Cisco storage networking solutions, including the Dell PowerVault MD3000i iSCSI Storage system. The Dell PowerVault MD3000i is an iSCSI SAN that can consolidate up to sixteen hosts, expands to support up to 18TB of data. The PowerVault MD3000i system's modular expandability provides the flexibility to add capacity as needed. The storage array can house up to fifteen 3.5-inch serially-attached SCSI (SAS) disk drives. Its capacity can be easily expanded by adding up to two PowerVault MD1000 expansion enclosures, for a total of 45 disk drives. For high availability, the PowerVault MD3000i supports redundant active/active controllers, management ports and power/cooling to provide real security at the hardware level and can automatically rebuild a failed drive using a global hotspare drive.

The server room has a fire suppression system, dedicated cooling, an uninterruptible power supply good for over six hours, and separate air filtering. Production servers running critical applications (eRoom^{TM}, databases, etc.) are clustered and configured for failover events. Servers are backed up through a dedicated backup server that connects across an internal Gigabyte network to a robotic tape drive. Incremental backups occur hourly Monday through Friday from 6am to 6pm. Incremental backups also are performed each night with full system backups occurring every Friday. Tapes are stored in a fireproof safe inside the server room, and full backups are taken off site on a weekly basis to a commercial storage facility.

8.4 Data Security

The Data Coordinating Center at the University of Utah has a dedicated, locked server room within its offices, and the building has 24 hour on-site security guards. The DCC has a state-of-the-art computer infrastructure and coordinates its network infrastructure and security with the Health Sciences Campus (HSC) information systems at the University of Utah. This provides the DCC with effective firewall hardware, automatic network intrusion detection, and the expertise of dedicated security experts working at the University. Network equipment includes three high-speed switches and two hubs. User authentication is centralized with two Windows 2003 domain servers. Communication over public networks is encrypted with virtual point-to-point sessions using secure socket layer (SSL) or virtual private network (VPN) technologies, both of which provide at least 128 bit encryption. The DCC will prepare an electronic data capture (EDC) system using commercial or open source products, and eRoom^{TM} is used for communications about the study. The EDC, eRoom^{TM} and other web applications use the SSL protocol to transmit data securely over the Internet.

Direct access to DCC machines is only available while physically located inside the DCC offices, or via a VPN client. All network traffic is monitored for intrusion attempts, security scans are regularly run against our servers, and our IT staff are notified of intrusion alerts. Security is maintained with Windows 2003 user/group domain-level security. Users are required to change their passwords every 90 days, and workstations time out after 10 minutes of inactivity. All files are protected at group and user levels; database security is handled in a similar manner with group level access to databases, tables, and views in Microsoft SQL Server.

8.5 Data Confidentiality

Prior to authorization to access Data Coordinating Center computer systems, all network investigators and research staff undergo training and sign confidentiality agreements. The investigators and staff of the data coordinating center are fully committed to the security and confidentiality of data collected for the study. All personnel at the DCC have signed confidentiality agreements concerning all data encountered in the DCC. Violation of these agreements may result in termination from employment at the University

of Utah. In addition, all personnel involved with DCC data systems have received Human Subjects Protection and HIPAA education.

8.6 Data Sharing Plan

It is a Federal requirement to share research data in a project such as this one. When the study is completed, the DCC will prepare a distributable database in compliance with these Federal requirements. This database will be de-identified sufficiently that it will not be subject to 45 CFR §46 nor the Health Insurance Portability and Accountability Act (HIPAA). Note that the limited PHI in the analysis database (month, year) is used in order to direct performance measure reports, and the month and year of visit will not be included in the database produced for sharing with other researchers.

Access to this research database will be managed in accordance with PECARN data sharing policies and applicable Federal laws. Currently this involves a signed research data use agreement between the investigator requesting the data and the University of Utah (PECARN DCC). This procedure is subject to alteration without requiring revision of this protocol document.

Registry data in this study will not be provided back to participating study sites except as outlined in the protocol. Each site, of course, has its own original data and can conduct its own quality improvement activities. However, the registry data will be de-identified and it will not be permissible for any investigator to attempt to re-identify data within the registry.

9 Protection of Human Subjects

9.1 Institutional Review Board Approval

Institutional Review Board (IRB) approval will be required from all sites participating in this study, including the PECARN Data Coordinating Center (DCC). The DCC will maintain documentation of initial and on-going approval at each site, and this documentation will be required in order for a site to submit electronic data to the DCC.

9.1.1 Expedited IRB Approval

Expedited IRB approval is requested for this study. Expedited IRB approval procedures are permissible when the research activities present no more than minimal risk to human subjects, and only involve procedures listed in specific categories listed in the regulations at Federal Register 63:29748, 1998, 21 CFR §56.110, and 45 CFR §46.110. For this project, the procedures fall under Category 5, which is research that involves materials (data, documents, records) that have been collected, or will be collected, solely for non-research purposes (e.g., medical treatment, diagnosis). All data in this study are from the medical records produced during the on-going clinical activities of the participating centers.

9.1.2 Waiver of Informed Consent and Assent

Waiver of informed consent and patient assent is requested for this study. As described in Section 5 on page 12, the participating sites and their component clinicians are the units of analyses in this study. The topic of study is the aggregate behavior of the emergency departments and potential changes in clinician performance measures, and selective consenting or assenting would destroy the scientific validity of the study. Performance feedback, while studied as part of this proposal, will be directed at quality improvement. Any practitioner or site specific feedback will be directed to that individual, identified only by a randomly generated study provider number, or site so that there is minimal risk of adverse financial or professional consequences during the study. Feedback will be provided so that the clinicians can better understand their own practice as a starting point for practice improvement. Practitioners will have the option of not reviewing provided feedback.

Waiver of informed consent from the approximately 300 clinicians is requested because risks to the individual clinicians are minimal (Section 9.2.3 on page 37), the waiver will not adversely affect the rights and welfare of the clinicians, uniform participation is necessary for the scientific validity of the study, and the clinicians will receive information about their performance during the study. These are the four requirements for waiver of informed consent outlined in 45 CFR §46.116(d).

Waiver of informed consent and assent from patients is requested for this project. Specific patients are not the subjects of study, but rather, their ED visit contributes information about the sites and clinicians, which are

the subjects of study. The requirements of 45 CFR §46.116(d) are also met for the patient population, as the risk is minimal, the rights and welfare of the patients are not being affected, and uniform inclusion of all patients is necessary for the scientific validity of the study. There is no provision for providing additional pertinent information to patients, as they are not the subjects of the study.

9.1.3 Waiver of Written Authorization for HIPAA

Waiver of authorization for use and/or disclosure of protected health information under the the Privacy Rule is requested for this study. Waiver of authorization for the use of patient electronic health record data is permissible under §164.512(i)(2)(ii). The requirements are that the use of the data involves no more than minimal risk to the privacy of the patients, that the research could not practicably conducted without the waiver, and the research cannot be conducted without access to the data. Written authorization is not practicable in the urgent environment of the ED, data are being abstracted long after the patient has departed the ED, and the scientific validity of the study would be destroyed since all visits must be included in order to determine the performance measures. The use or disclosure of the protected health information involves no more than minimal risk to the individuals because:

- There is an adequate plan to protect the identifiers from improper use and disclosure. The PECARN DCC has provided Business Associate Agreements for execution by each participating site, and the security of its information systems has been described in Section 8.4 on page 32.
- There is provision to destroy identifiers before integrating patient data into the registry, and the de-identification procedures are thoroughly discussed in Section 6.2 on page 18.
- The protected health information will not be reused or disclosed to any other persons or entities, except as required by law, or for authorized oversight of the research project.

9.2 Risks to Human Subjects

9.2.1 Research Participants and Characteristics

Number and characteristics of participants. This project will develop a registry of electronic health record data including all visits to the base and

satellite emergency departments (EDs) of the four sites for all patients. The registry will contain data from the electronic health record from all visits to the EDs from CY 2012 and during an 24 month study period between 2012 and 2015. There will be (estimated) over 1,380,000 visits during that time and 16% will involve children less than 1 year old, 37% between the ages of 1 and 4 years, 20% between 5 and 9 years old, 16% between 10-14 years old, and 10% between 14-18 years of age. The children will be African-American (46%), Caucasian (44%), Asian, Native Hawaiian or Other Pacific Islander (1.4%), American Indian or Alaskan Native (0.2%) and 14% of Hispanic ethnicity.

There will be over 300 licensed independent practitioners caring for the children involved in the visits in the registry. Eighty-seven percent are physicians, 40% board certified in pediatric emergency medicine and 45% in pediatrics. They have been in practice for a range of years (47% $<$ 5 years, 26% 5-10 years, and 27% $>$10 years).

Inclusion in the registry will be regardless of race, ethnicity, or gender. Inclusion will be regardless of diagnosis or chronic health condition. Clinicians will be included in the protocol regardless of race, ethnicity, or gender.

Collaborating sites and roles. There will be four sites contributing data to the registry. Each of these sites will contribute data from the both a base and satellite ED. The Data Coordinating Center for this project is at the University of Utah.

9.2.2 Sources of Materials

This project includes only data that will be obtained from already completed EHRs. No prospective data will be collected solely for study purposes. Queries of the Pediatric Emergency Care Registry will provide the data for analysis. At each site, the investigators involved in the project, the site research coordinator and information technology contact will have access to identifiable private information about some or all subjects. After de-identification procedures are finalized by the DCC (Section 6.2 on page 18), the procedures will be carried out at each clinical site. Then components of the EHR will be submitted electronically using a secure system hosted by the DCC. At the DCC, the DCC PI, data managers, statisticians, and other staff may have access to a limited set of identifiable information.

9.2.3 Potential Risks

This study is without direct patient contact and utilizes only existing electronic health record data. Data will be acquired from hospital computer systems and the primary potential risk to subjects is improper disclosure of medical information. This risk is minimized by data management steps outlined above.

The performance measures involved in the study have been carefully developed to include only ones with a strong evidence base and endorsement by nationally recognized experts and stakeholders. We expect that the quality of clinical care will only improve with the intervention. Outcomes will be followed on a monthly basis to insure that no unexpected decline in clinical care is observed. Should this very unlikely event occur, the study team would intervene to address the problem including, if needed, discontinuing the intervention. Risk to the participating sites or practitioner is the disclosure that any one site or practitioner has quality outcome patterns that greatly differ. This risk will be mitigated by use of deidentified practitioner and site study numbers assigned by the DCC for the Registry and report cards. The DCC will maintain a study key to allow for appropriate identification of visits pertaining to each practitioner and also to allow for distribution of the report cards. This will decrease the risk to practitioners as identification is not known by the site administration. All feedback reports will use codes to protect the identification of sites and clinicians, and all published results will be presented in aggregate form.

9.3 Adequacy of Protection Against Risks

This project will entail the development of an emergency department electronic health record registry and evaluation of a quality of care measures from this registry from four sites for all patients. All data will be obtained from already completed electronic health records. We will collect no prospective data solely for study purposes. Data analysis will be conducted retrospectively using de-identified data and will involve a comparison of the performance measures using a staggered time-series analysis. In every case, final decisions regarding clinical care will have been made between patients, their parents, and clinicians at the bedside, prior to the transmittal of data to the DCC or provision of performance measure information to sites and clinicians. This intervention will in no way mandate clinicians to pursue a specific treatment for a given patient. However, clinicians may be better

informed about their own prior performance and will have an opportunity to provide evidence-based improvements in the quality of care they provide to future patients.

9.3.1 Protection Against Risks

Data will be acquired from hospital computer systems and the primary potential risk to subjects is improper disclosure of medical information. Data analysis for this project will be conducted using de-identified data. Data will be housed at the DCC, for which security was previously described (Section 8.4 on page 32).

The performance measures involved in the study have been carefully developed to include only ones with a strong evidence base and endorsement by nationally recognized experts and stakeholders. We expect that the quality of clinical care will only improve with the intervention. Outcomes will be followed on a monthly basis to insure that no unexpected decline in clinical care is observed. Should this very unlikely event occur, the study team would intervene to address the problem including, if needed, discontinuing the intervention. All published results will be presented in the aggregate and care will be taken to keep identification of sites and investigators confidential.

9.3.2 Human Subjects Protection Training

All key personnel have completed Human Subjects Protection Training as mandated by involved Institutional Review Boards. This training involves an extensive, web-based, curriculum emphasizing the safeguards necessary to conduct human subjects research, particularly with children.

9.4 Potential Benefit of Research

This project will compare quality performance measures of emergency healthcare for children across eight EDs (base and satellite) in the Pediatric Emergency Care Applied Research Network. It will compare quality performance measures of emergency care provided to children across different institutions by using data extracted from electronic health records. It will also establish benchmark levels for performance measures using qualitative and quantitative methods. Once these have been delineated it will implement an intervention of feedback to involved institutions and practitioners and measure the impact of the feedback protocol on quality measures of care. It will evaluate the impact of providing provider-specific feedback and

benchmarks of care on quality measures that are within the locus of control of the individual provider. The feedback may directly improve the quality of care of future patients treated by the practitioners and the sites.

9.5 Importance of Knowledge to be Gained

In its 2006 report, "Emergency Care for Children: Growing Pains", the Institute of Medicine recommended that pediatric emergency medical systems support the development and measurement of standards for emergency care performance measurement. The ability to accurately and comprehensively assess the process and outcomes of care in pediatric emergency patients is imperative to the evaluation of the quality of care provided. Past endeavors have been limited by the labor intensive nature of obtaining the information needed to determine quality performance measures. This project will allow for the comprehensive and scalable determination of quality performance measures using an electronic health record registry. This project will set benchmarks and targets needed to improve processes and patient outcomes and will compare quality of PEM care across sites and providers. It will evaluate the impact of providing provider-specific feedback and benchmarks of care on quality measures that are within the locus of control of the individual provider. The feedback may directly improve the quality of care of future patients treated by the practitioners. In addition, the systematic and widespread collection and reporting of performance and outcomes, using the same operational definitions, is critical to allow clinicians and other emergency care stakeholders to work together to innovate to improve care beyond the local level. The project thus has enormous potential to improve our ability to evaluate EMSC systems of health care delivery, as well as to lead to improvements in the quality of care provided to acutely ill or injured children.

10 Health Insurance Portability and Accountability Act

All relevant data in the electronic health record is being collected in this study, including potentially identifying information, because the entire project aims to create a registry automatically from the medical record. Rigorous de-identification procedures will be conducted to produce the analysis database with a minimum of Protected Health Information (PHI). Specifically, the analysis database will have the month and year of visit. If unexpected addi-

tional PHI is detected in submitted data by the Data Coordinating Center, it will be de-identified by the DCC, and feedback will be provided to the clinical site(s) to further refine the de-identification procedures so that this does not recur.

Data elements for race, ethnicity, and gender of subjects will be preserved. These demographic data are required for Federal reporting purposes to delineate subject accrual by race, ethnicity, and gender.

All study sites have been offered a Business Associate Agreement with the University of Utah, which agrees to handle submitted data with the security precautions required for PHI. Copies of executed Business Associate Agreements are maintained at the DCC.

11 Inclusion of Women and Minorities

Data from the emergency department (ED) electronic health record of all visits for patients (both male and female) seen at the protocol sites will be included. Although patients will be enrolled regardless of gender, we anticipate a small preponderance of male patients (52-55%) reflecting the distribution of patients presenting to the participating EDs.

Data from the emergency department (ED) electronic health record of all patients seen at the protocol sites will be included. Visits will be included without regard to patient race or ethnicity. A diversity of ethnic and racial backgrounds is represented at the selected PECARN sites. The overall distribution of the patient population from these sites is as follows: African-American (46%), Caucasian (44%), Asian/Pacific Islander (1.4%), American Indian or Alaskan Native (0.2%). Fourteen percent are of Hispanic ethnicity.

12 Inclusion of Children

Data from the emergency department electronic health record for all visits of patients will be collected from the selected sites. Assessment of quality performance measures for pediatric emergency care is essential for all children regardless of age. ED visits will be included without regard to age. Members of the investigative team are board certified in pediatrics and pediatric

emergency medicine and have extensive experience and expertise in caring for children in this age range. The sites participating all have emergency departments dedicated to the care of children.

13 Access to and Retention of Records

For federally funded studies subject to the Common Rule, records relating to the research conducted shall be retained for at least 3 years after completion of the research. Completion of the research for this protocol should be anticipated to include planned primary and secondary analyses, as well as subsequent derivative analyses. Completion of the research also entails completion of all publications relating to the research. All records shall be accessible for inspection and copying by authorized representatives of the regulatory authorities at reasonable times and in a reasonable manner [45 CFR §46.115(b)].

Bibliography

[1] Health, United States, 2005: Chartbook on Trends in the Health of Americans. Hyattsville, Maryland: National Center for Health Statistics; 2005.

[2] Health, United States, 2009: With Special Feature on Medical Technology. Hyattsville, Maryland: National Center for Health Statistics; 2010.

[3] Institute of Medicine Committee on the Future of Emergency Care in the US Health Care System. Emergency Care for Children: Growing Pains. Washington, DC: National Academy Press; 2006.

[4] Seidel JS, Henderson D, Tittle S, Jaffe D, Spaite D, Dean JM, et al. Priorities for research in emergency medical services for children: results of a consensus conference. Pediatr Emerg Care. 1999 Feb;15(1):55–8.

[5] Chamberlain JM, Patel KM, Pollack MM. Association of emergency department care factors with admission and discharge decisions for pediatric patients. J Pediatr. 2006 Nov;149(5):644–649.

[6] Crain EF, Weiss KB, Fagan MJ. Pediatric asthma care in US emergency departments. Current practice in the context of the National Institutes of Health guidelines. Arch Pediatr Adolesc Med. 1995 Aug;149(8):893–901.

[7] Crain EF, Mortimer KM, Bauman LJ, Kercsmar CM, Weiss KB, Wissow L, et al. Pediatric asthma care in the emergency department: measuring the quality of history-taking and discharge planning. J Asthma. 1999;36(1):129–38.

[8] Hampers LC, Trainor JL, Listernick R, Eddy JJ, Thompson DA, Sloan EP, et al. Setting-based practice variation in the management of simple febrile seizure. Acad Emerg Med. 2000 Jan;7(1):21–7.

[9] Ramenofsky ML, Luterman A, Quindlen E, Riddick L, Curreri PW. Maximum survival in pediatric trauma: the ideal system. J Trauma. 1984 Sep;24(9):818–23.

[10] Stanley RM, Teach SJ, Mann NC, Alpern ER, Gerardi MJ, Mahajan PV, et al. Variation in ancillary testing among pediatric asthma patients seen in emergency departments. Acad Emerg Med. 2007 Jun;14(6):532–8.

[11] Zerr DM, Del Beccaro MA, Cummings P. Predictors of physician compliance with a published guideline on management of febrile infants. Pediatr Infect Dis J. 1999 Mar;18(3):232–8.

[12] Cain VS, Kington RS. Investigating the role of racial/ethnic bias in health outcomes. Am J Public Health. 2003 Feb;93(2):191–2.

[13] Pediatric Emergency Care Applied Research Network. The Pediatric Emergency Care Applied Research Network (PECARN): rationale, development, and first steps. Acad Emerg Med. 2003 Jun;10(6):661–8.

[14] Pediatric Emergency Care Applied Research Network. The Pediatric Emergency Care Applied Research Network (PECARN): rationale, development, and first steps. Pediatr Emerg Care. 2003 Jun;19(3):185–93.

[15] Alpern ER, Stanley RM, Gorelick MH, Donaldson A, Knight S, Teach SJ, et al. Epidemiology of a pediatric emergency medicine research network: the PECARN Core Data Project. Pediatr Emerg Care. 2006 Oct;22(10):689–99.

[16] Alpern ER, Donaldson AE, Alessandrini EA, Gorelick MH, Stanley RM, Teach SJ, et al.; PAS. Epidemiology of pediatric emergency department recurrent visits. PAS. 2005;57:1862.

[17] Mahajan P, Alpern ER, Grupp-Phelan J, Chamberlain J, Dong L, Holubkov R, et al. Epidemiology of psychiatric-related visits to emergency

departments in a multicenter collaborative research pediatric network. Pediatr Emerg Care. 2009 Nov;25(11):715–20.

[18] Timm N, McAneny C, Alpern E, Macy M, Ruddy RM. Is pediatric emergency department utilization by pregnant teens on the rise? PAS. 2010;p. 20101494.434.

[19] Gorelick MH, Alpern ER, Singh T, Snowdon D, Holubkov R, Dean JM, et al. Availability of pediatric emergency visit data from existing data sources. Acad Emerg Med. 2005 Dec;12(12):1195–200.

[20] Gorelick MH, Knight S, Alessandrini EA, Stanley RM, Chamberlain JM, Kuppermann N, et al. Lack of agreement in pediatric emergency department discharge diagnoses from clinical and administrative data sources. Acad Emerg Med. 2007 Jul;14(7):646–52.

[21] Alessandrini EA, Alpern ER, Chamberlain JM, Shea JA, Gorelick MH. A new diagnosis grouping system for child emergency department visits. Acad Emerg Med. 2010 Feb;17(2):204–13.

[22] Alessandrini EA, Alpern ER, Chamberlain JM, Gorelick MH. Developing a diagnosis-based severity classification system for use in emergency medical systems for children. PAS. 2007;p. 61670.5.

[23] Dean JM, Chamberlain J, Gorelick MH, Stanley R, Snowdon D, Mahajan P, et al. The PECARN Core Data Project: benchmarking hospitals in a pediatric emergency medicine research network. Pediatr Res. 2004;55:119A.

[24] Stanley R, Alpern ER, Gerardi M, Mahajan P, Teach SJ, Chamberlain JM, et al. Variations in diagnostic testing in the ED for pediatric non-urgent illnesses. PAS. 2005;57:2865.

[25] Gorelick MH, Stevens MW, Schultz TR, Scribano PV. Performance of a novel clinical score, the Pediatric Asthma Severity Score (PASS), in the evaluation of acute asthma. Acad Emerg Med. 2004 Jan;11(1):10–8.

[26] Kiefe CI, Weissman NW, Allison JJ, Farmer R, Weaver M, Williams OD. Identifying achievable benchmarks of care: concepts and methodology. Int J Qual Health Care. 1998 Oct;10(5):443–7.

[27] Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving quality improvement using achievable benchmarks

for physician feedback: a randomized controlled trial. JAMA. 2001 Jun;285(22):2871–9.

[28] Weissman NW, Allison JJ, Kiefe CI, Farmer RM, Weaver MT, Williams OD, et al. Achievable benchmarks of care: the ABCs of benchmarking. J Eval Clin Pract. 1999 Aug;5(3):269–81.

[29] Neamatullah I, Douglass MM, Lehman LwH, Reisner A, Villarroel M, Long WJ, et al. Automated de-identification of free-text medical records. BMC Med Inform Decis Mak. 2008;8:32.

[30] Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman LW, Moody G, et al. Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database. Crit Care Med. 2011 May;39(5):952–60.

[31] Friedlin FJ, McDonald CJ. A software tool for removing patient identifying information from clinical documents. J Am Med Inform Assoc. 2008;15(5):601–10.

[32] Meystre SM, Friedlin FJ, South BR, Shen S, Samore MH. Automatic de-identification of textual documents in the electronic health record: a review of recent research. BMC Med Res Methodol. 2010;10:70.

[33] Uzuner O, Sibanda TC, Luo Y, Szolovits P. A de-identifier for medical discharge summaries. Artif Intell Med. 2008 Jan;42(1):13–35.

[34] Uzuner O, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. J Am Med Inform Assoc. 2007;14(5):550–63.

[35] Crowley RS, Castine M, Mitchell K, Chavan G, McSherry T, Feldman M. caTIES: a grid based system for coding and retrieval of surgical pathology reports and tissue specimens in support of translational research. J Am Med Inform Assoc. 2010;17(3):253–64.

[36] Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform. 2008;p. 128–44.

[37] Uzuner O. Second i2b2 workshop on natural language processing challenges for clinical records. AMIA Annu Symp Proc. 2008;p. 1252–3.

[38] Zweigenbaum P, Demner-Fushman D, Yu H, Cohen KB. Frontiers of biomedical text mining: current progress. Brief Bioinform. 2007 Sep;8(5):358–75.

[39] Montgomery DC. Introduction to Statistical Quality Control. 6th ed. Hoboken, NJ: John Wiley and Sons; 2009.

[40] Little RJA, Rubin DB. Statistical Analysis with Missing Data. 2nd ed. Hoboken, NJ: Wiley-Interscience; 2002.

[41] Rubin DB. Inference and missing data. Biometrika. 1976;63:581–92.

[42] Shafer JL. Analysis of Incomplete Multivariate Data. London: Chapman & Hall; 1997.

[43] Shewhart WA. Finding causes of quality variations. Manufacturing Industry. 1926;11:125–128.

[44] Western Electric. Statistical Quality Control Handbook. Indianapolis, IN: AT&T; 1956.