FINAL REPORT

PEAT: Pediatric Emergency Assessment Tool

Principal Investigator: Marc H. Gorelick, MD, MSCE
Team Members: Kathleen Cronan, MD
Justine Shults, PhD
Jo Bergholte, MS

Organization: Medical College of Wisconsin

PI Contact Information: Children's Hospital of Wisconsin MS#677
9000 W. Wisconsin Ave.
Milwaukee, WI 53226
(414) 266-2625
(414) 266-2635 (fax)
mgorelic@mail.mcw.edu

Dates of Project: 9/30/2000-9/29/2003

Federal Project Officer: Dr. Marlene Miller (previously Dr. Elinor Walker)

This project was supported by grant R03 HS11395 from the Agency for Healthcare Research and Quality.

1. ABSTRACT

Purpose: To develop and validate a multivariable model, using predictor information available at the time of patient triage, to predict the probability of the need for different levels of care among pediatric emergency patients.
Scope: Evaluations of quality and outcomes of care in emergency medical services for children (EMSC), and comparisons of outcomes between EMSC systems or components of systems, require a method of adjusting for important differences in acuity between patient populations. No satisfactory method of risk adjustment currently exists that is broadly applicable to all EMSC patients.
Methods: Retrospective cohort study of 5500 children age 18 years and under treated at four emergency departments over a 12-month period. Data were obtained from abstraction of patient records. Logistic regression was used to develop models to predict receipt of any non-routine care in the ED, and admission to the hospital. Data on ED length of stay and hospital charges were also obtained.
Results: Eight predictor variables were included in the final models: presenting complaint; age; triage acuity category; arrival by emergency medical services; current use of prescription medications; and triage vital signs (heart rate, respiratory rate, temperature). The resulting models had excellent calibration and discrimination in both derivation and validation samples. The Revised Pediatric Emergency Assessment Tool (RePEAT) score then was calculated as the sum of the predicted probability of receiving care and twice the predicted probability of admission. The RePEAT score was significantly associated with ED charges and with ED length of stay, and it contributed significantly to models comparing these outcomes across sites, providing a useful measure for risk adjustment.
key words: risk adjustment, emergency medical services, pediatric

2. PURPOSE

The primary objective of this study was to develop and validate a predictive model to be used as a risk-adjustment tool for use in evaluating outcomes of pediatric emergency care. The model uses information readily available and routinely recorded at the time of triage to predict intensity of services required. This model may then be used to calculate an acuity score – the Revised Pediatric Emergency Assessment Tool (RePEAT) – that reflects the expected probabilities of each level of care for patients. These expected probabilities can then be used to adjust for differences in underlying acuity in evaluations of outcomes of care in EMSC.

The specific aims of the study were to:

1) Develop a clinically meaningful and analytically useful method for categorizing presenting complaints;

2) Develop and validate a predictive model, using multivariable regression, to predict intensity of resource use for pediatric emergency patients;

3) Evaluate the usefulness of this model in risk adjustment by assessing the degree to which the output of the model correlates with other clinically relevant outcomes: ED charges, and time to move a patient through the ED (ED throughput time).

3. SCOPE: BACKGROUND AND SIGNIFICANCE

A. Need for an acuity adjustment tool for EMSC. Emergency Medical Services (EMS) systems are a relatively new component of our national healthcare system.[1] Over the past 2 decades, attention has become focused on the unique needs of children within the EMS system, and increasing resources have been devoted to the establishment of various components of Emergency Medical Services for Children (EMSC). To date, much of the research on EMSC has been descriptive. [2] However, in more recent years, the need for critical evaluation of the various components of EMSC has been recognized.[3] Among the areas in need of evaluation are the effectiveness of different components of EMSC, outcomes of different configurations of EMSC, resource allocation and utilization both within and between emergency medical systems, and cost-effectiveness of EMSC and its components. The Institute of Medicine's (IOM) Committee on Emergency Medical Services for Children, in its 1993 report, called for greater research on the effectiveness of EMSC and its various components.[3] Areas singled out for particular emphasis included comparisons of outcomes of different configurations of EMSC and of its various components as well as, economic evaluations including cost-effectiveness studies.

Such studies frequently are not well suited to randomized trials and instead are conducted using observational study methods. A major potential limitation of such nonexperimental designs is confounding, specifically, in this case, by the types and severity of illness of patients treated in different settings or systems.[1,4] It is widely recognized in health services research that comparisons of outcomes and resource utilization must take into account important differences in severity or acuity in the patient populations being compared.[4] Measures for case-mix or risk adjustment are available for general hospital, critical care, and trauma patients.[5,6,7,8] However, no broadly applicable instrument for risk adjustment is yet available for emergency medicine. This has severely hindered the progress of systems evaluation and health services research in the field.[2] Indeed, the development of a general acuity measure for children was named as one of the top priority areas for research in the Institute of Medicine report on EMSC.[3]

B. Requirements for an acuity index for EMSC. The desirable characteristics of a risk adjustment tool have been described.[7,9,10,11] These include relative simplicity, clinical credibility, data availability, and reliability. Perhaps most important, a risk adjustment index must be **valid**; that is, it must accurately predict clinically relevant outcomes of interest. The index should have a high, known correlation with the patient's underlying severity of illness, usually measured as mortality or morbidity. This correlation should be consistent for dissimilar patients in a variety of settings.

Importantly, a risk adjustment index should be applicable to all pediatric emergency patients, not just the minority of such patients that are seen in pediatric emergency departments. The index should also make clinical sense (i.e., have face validity). For EMSC, an important outcome is the **need** for emergency medical services. As discussed in more detail below, need for services is related to a number of factors, both measurable and unmeasurable, including severity of illness and access to alternative sources of care. For this study, the key outcome variable is the amount of care provided to the patient in the emergency department. This outcome was selected because it reflects whether, in the judgment of the physician who treated the patient, the patient needed to be seen.

A second important property is **reliability**: an index should have clear and objective rules for deriving the score, so that the same rater over time or different raters at the same time will derive an identical score for the same case. Finally, to be useful, an index must be **practical**. The data required for derivation of the acuity score must be readily available for all patients for whom the index is intended, routinely recorded, and easily applied by nonclinicians. The score should also be as simple and parsimonious as possible.

Two of these requirements for a useful acuity measure pose a particular challenge for EMSC. The first is the issue of data availability. Many of the existing severity assessment instruments in existence (discussed in the following section) rely on detailed clinical information obtained during the course of the patient's care. Such information can often be obtained only through comprehensive chart review, diminishing its utility for research in health services and systems evaluation. More importantly, the data in some of these indices, such as laboratory studies, are not obtained for the majority of patients who access EMSC. To be broadly applicable, an acuity measure for EMSC should be limited to data that are routinely obtained and recorded for all patients. The IOM report recommends the development of a uniform data set for emergency medical services as a means of improving research in the field.[3] Documentation of many of these elements is already mandated by The Joint Commission.[12] In the proposed study, only variables included in this list or The Joint Commission requirements will be studied.

C. Definition of severity. Second, we must specify what is meant by severity. Severity of illness is a commonly used concept in medicine, but it is often difficult to define.[13] It is an inherent characteristic of an illness that reflects the natural history, that is, the prognosis in the absence of intervention.[14] As such, it is closely related to the concept of acuity, or the need for care; patients with a higher level of severity require greater levels of care and require care more urgently. For purposes of the current study, the terms severity and acuity will be used interchangeably. Both are indications of the patient's physiologic state, without regard to treatment available or rendered.

Severity is not generally directly measurable, because patients are not usually observed without intervention. In the absence of a gold standard, severity must be inferred from a related measure. In one commonly used method, which we term an outcome based approach, severity is related to the likelihood of an observable outcome such as mortality,[13,15,16,17] risk of organ system failure,[18] or disability.[19] The underlying assumption is that the observed outcomes are reflective of patient severity, even though the outcome is also influenced by the care received.[20]

An alternative approach, termed resource based, relates severity to the level of care required.[21] Examples include need for interventions on transport,[22] emergency surgery,[23] or hospital admission.[24,25] The Therapeutic Intervention Scoring System (TISS) is a severity index originally designed for the intensive care setting that quantifies the number and intensity of therapeutic procedures performed.[26] Most of the severity indices that use the resource-based approach are based at least implicitly on the assumption that only necessary services will be rendered; therefore services provided are a reasonable proxy for services required.[23,24,26] Some investigators have questioned this assumption, arguing that variability between practitioners -- practice style -- precludes using actual care as a measure of care that is needed.[9] However, practice variation reflects only in part differences in the standard of care provided but also acceptable alternatives within the standard of care.[27,28]

When averaged over a large number of physicians, the care provided to patients is likely to reflect an average, reasonable judgment that such care was necessary. In fact, severity scores based on risk of mortality have been shown to correlate extremely well with resources used (e.g., TISS score), indicating reasonable construct validity of the resource based approach to measuring severity.[11,29]

The choice of which approach to take in developing a tool for EMSC must account for several factors unique to the pediatric emergency medicine context. First, serious adverse outcomes, such as mortality or severe disability, are rare. Second, if severity is a reflection of outcome in the absence of an intervention, diagnostic as well as therapeutic interventions must be considered, because both types of interventions affect the outcome of an EMSC encounter. Indeed, evaluation is a primary function of EMSC.[30] A third, related, consideration is the element of temporality. Many severity indices are based largely on the patient's diagnosis. This may be appropriate for hospital inpatients, ICU patients, or trauma patients, in whom at least a provisional diagnosis is usually known at the start of the period of care in question. In the emergency medicine setting, however, a diagnosis is not typically made until the encounter is well underway. The diagnosis, in fact, reflects at least in part the results of the initial evaluation and treatment. A neonate with high fever has the same need for evaluation whether the final diagnosis is viral upper respiratory infection or meningitis. If a distinction cannot be made after initial evaluation, the child may well have the same need for treatment. The need for care (diagnosis or treatment), therefore, should properly be determined based on information available before care is actually rendered.[31] Last, a risk adjustment tool appropriate for use in EMSC must also take into account heterogeneity in patients and illnesses, which may preclude the use of a single outcome measure. For example, a child with an isolated facial laceration has a high risk of morbidity in the form of poor future cosmetic results in the absence of treatment but has an extremely low risk of mortality or of requiring hospital admission.

The rarity of severely adverse outcomes and heterogeneity of outcomes of interest make the outcome-based approach to defining severity, in which patients are stratified according to risk of, for example, mortality, unlikely to provide good discriminative ability. The resource-based approach -- in which intensity of services required is used as the measure of severity or acuity -- is better suited to the considerations listed in the preceding paragraph.

For the purposes of this study, then, severity refers to the need for emergent evaluation and treatment, as best it can be determined before the patient is evaluated and treated by the emergency physician. In this sense, severity reflects the probability that the patient would have an adverse outcome were he not treated in the ED. The ideal severity adjustment instrument would reflect the decision that a capable physician or nurse, presented with the information available prior to evaluation by the emergency physician, would make about the urgency with which each patient needed to be seen. To reflect the information available to the decision maker, the severity adjustment instrument must require no more data than can be obtained by a triage nurse (i.e., vital signs and a brief interview). Use of triage data for classification of severity is also necessary from a pragmatic perspective. Complete information from a clinical evaluation will not be available for patients who are denied authorization to be seen in the ED or who leave without being seen by a physician.

D. Severity adjustment tools available and their limitations. A large number of instruments has been developed to adjust for severity or case-mix in clinical research and quality improvement activities.[5,6,7,9,11] Several measures are designed for hospital inpatients. Examples include MEDISGROUPS®, the Computerized Severity Index, and the Patient Management Categories (PMC) system.[6] Two severity indices were developed for patients in intensive care units: the APACHE[32] system for adults and the PRISM (Pediatric Risk of Mortality) score for pediatric patients.[13,14] Although several of these indices perform well in other contexts, they are unsuited for use in EMSC for several reasons: the use of risk of mortality as an outcome, the need for detailed clinical information, and reliance on retrospective determination of severity.

The limitations to employing these measures in the EMSC context are illustrated by the poor predictive ability of the PRISM score when applied to interhospital transport.[33,34]

Several disease-specific acuity scoring systems are available for use in EMSC. The majority of these are intended for use in trauma patients, including the Injury Severity Score, Trauma Score, and Pediatric Trauma Score.[35] Others are available for asthma,[36] croup,[37] and febrile young children.[38] These measures, however, cannot be applied broadly to all patients. A tool has been developed specifically for measuring severity of patients undergoing interhospital transport,[22] but such an instrument also has limited applicability. More global EMSC evaluations, for example, comparing performance of different types of EDs, require risk adjustment tools that can be applied to all types of patients.

One potential approach is to use triage status as a measure of severity. Triage is the assignment of patient priority, the process of assigning the order in which patients are seen in the ED.[39] Many different methods of triage are used. Typically, a nurse briefly interviews the patient (or parent) on arrival and obtains vital signs, and she or he then uses this information to determine the priority with which the patient should be seen. The usual classification scheme divides visits into three levels of increasing priority: nonurgent (or nonacute), urgent (or acute), and emergent. Most hospitals have formal triage criteria to guide the assignment of a priority level, and some systems of triage use formal algorithms, including computer-based systems.[40,41,42] Although the triage score meets many of the requirements for a severity adjustment tool enumerated above, there are important drawbacks. The most important is that reliability of triage assignment even within an institution has been shown to be relatively poor, probably reflecting differences in the methods of triage and the incorporation of subjective judgments by the triage personnel into the assignment.[24,43,44,45] Variability of triage methods between EDs limits the usefulness of assigned triage level for evaluation and comparison of different EDs. An additional limitation to triage level is that visits are classified into a small number of categories -- typically two or three -- that may be insufficient for risk stratification.

A potential risk-adjustment tool specific to pediatric emergency medicine has recently been proposed, the Pediatric Risk of Admission (PRISA) score.[25,46] This score has been shown to predict accurately the risk of admission to the hospital of patient seen in a pediatric ED. It has a number of limitations, however. First, because it was initially developed at a single institution, a subsequent validation study showed the need for recalibration of the model. Second, predictor variables included not only data collected at triage but also results of laboratory and radiographic examinations (many of which were not measured in the large majority of patients). As noted above, this also limits the applicability of this tool. Finally, only hospital admission is considered as an important outcome, the limitations of which are outlined above. We believe our study overcomes many of these limitations and provides an opportunity to develop a more suitable risk-adjustment measure.

E. Applications of the RePEAT score. There are a number of potential applications of such a predictive model. The primary intended use is in the evaluation of EMSC systems and their components. Acuity scores can be used to adjust for patient mix when comparing outcomes between EDs or between prehospital care systems. Examples of outcomes for which such adjustment is important include patient care costs, rates of triage away from the ED, ED throughput time, and relapse or revisit rates. Increasingly, providers (including hospitals and physicians) are being evaluated and ranked on the basis of resource utilization, with important implications for compensation. It is essential to know what the expected utilization is and correct for it for such comparisons to be valid. In addition, individual EDs can compare actual resource use with that expected from the model in internal quality improvement activities and in resource allocation and planning.

We recognize that our proposed predictive model, based as it is on relatively limited information, may fail to perform well in correctly categorizing the outcome of a given individual patient. It is worthwhile to emphasize that the RePEAT is intended to be applied to groups of patients, such as those seen in a given ED or during a particular time period.

We do not envision the RePEAT being used to make decisions about individual patients either prospectively (e.g., for triage), or retrospectively (e.g., for approval or denial of ED care).

F. Summary. The IOM report, recognizing the lack of an appropriate general acuity measure for ill or injured children, named the development of such an instrument as one of the top priorities for research in EMSC.[3] The proposed study will address this critically important need. We intend to develop a broadly applicable index of acuity for pediatric emergency patients that will validly and reliably predict the level of services required. Scoring will use readily available data routinely obtained at the initiation of the visit, enhancing the practicality of the proposed instrument. The resulting acuity score, the RePEAT, will permit risk adjustment when making evaluations and comparisons of outcomes of emergency care. The proposed project thus has enormous potential to improve our ability to evaluate systems of EMSC as well as to lead to improvements in the quality of care provided to acutely ill or injured children.

4. METHODS

A. Study Design

This was a retrospective cohort study, in which a cohort of children presenting to an ED for care was followed through the course of the visit, and data on predictor (exposure) and outcome variables were collected. Data on both predictors and outcomes were abstracted from existing records after the outcome had occurred, making this a retrospective study.

B. Setting and Subjects

The study was conducted at the emergency departments of four hospitals. All children presenting to the EDs during a 12-month period were eligible for the study. A systematic sample of eligible visits, in which every *n*th visit was chosen, was selected at each of the participating hospitals. The sampling fraction was calculated differently at each hospital to provide the desired sample size.

Hospital	Location	Type of facility	Sampling fraction
Children's Hospital of Wisconsin	Milwaukee, WI (urban/suburban)	free-standing pediatric hospital	1 in 24
St. Mary's Hospital	Milwaukee, WI (urban)	general hospital with general ED	1 in 5
AI DuPont Hospital for Children	Wilmington, DE (suburban)	free-standing pediatric hospital	1 in 19
Beebe Medical Center	Lewes, DE (rural)	general hospital with general ED	1 in 4

C. Measurements.

i. Predictor variables. The predictor variables consist of reported and observed patient-level data, as well as circumstantial data, that are routinely obtained during the triage process. The reported data include age, race, sex, the reason for the visit, past medical history, current medication use, and means of transportation. The observed data include heart rate, respiratory rate, blood pressure (systolic and diastolic), temperature, oxygen saturation as measured by pulse oximetry, and mental status as measured by the Glasgow Coma Scale. Circumstantial data include time of day and month of the visit. These are summarized in Table 1.

Table 1. Potential Predictor Variables

Predictor variable	Definition for model
Age	3 groups: <3 mo., 3-24 mo., >24 mo.
Chief complaint (RVC code)	2 ordinal variables:
	1-5 based on quintile of receiving any care in ED
	1-5 based on quintile of admission
Past medical history	dichotomous based on report of qualifying condition
Current medications	dichotomous based on reported current use of prescription medications
Mode of transportation	dichotomous based on arrival via EMS, air or ground transport, or police vs. walk-in or self-transport
Triage status	3 levels: nonurgent, urgent, emergent
Heart rate	% of age appropriate normal value
Respiratory rate	% of age appropriate normal value
Temperature	continuous
Blood pressure	% of age appropriate normal value
Oxygen saturation	continuous (only included in presence of indication)
Glasgow Coma Scale	ordinal (only included in presence of indication)
Season	4 categories based on date of visit (winter, spring, summer, fall)
Time of day	3 categories based on time of arrival in ED: day (0800-1559), evening (1600-2359) and night (0000-0759)

The reason for the visit was categorized according to the Reason for Visit Classification for Ambulatory Care (RVC) of Schneider et al.[47] The RVC is a classification of lay terminology encountered in the ambulatory care setting. The emphasis of the RVC is on the patient's motivation for seeking medical care and her or his perspective of the problem. It has undergone extensive testing and revision and has been used for a number of health surveys, including the National Hospital Ambulatory Medical Care Survey (NHAMCS).[47,48] The RVC also includes an alphabetic index of terms with their appropriate code numbers for ease of use and instructions for the application of the coding system.

Although there are several hundred different RVC codes, data from the NHAMCS survey show that 70% of all ED visits (including adults and children) are classified in the symptom module and 21.6%, in the injuries and adverse effects module.[48] Moreover, the 20 most frequently mentioned principal reasons for visit accounted for almost half of all visits. The number of codes likely to be encountered is therefore relatively small. However, because even this restricted number of codes would be difficult to accommodate in a predictive model, further grouping of the data is necessary.

First, a system of clustering chief complaints called the Pediatric Emergency Reason for Visit Clusters, or PERC, was developed. Investigators began with a group of candidate clusters derived from the set of complaint groupings used in the GE Medical Systems ED Tracker® and other commercially available tracking systems as well as those included in the Pediatric Emergency Care Applied Research Network Core Data Project chief complaint list. Candidate clusters were first reviewed for clinical sensibility and were combined, when appropriate, and new ones were added if needed. Individual RVC codes were then assigned to the appropriate PERCs.

Data for further development of the PERCs were derived from the NHAMCS ED datasets for 1998 and 2000. All pediatric visits in the dataset were assigned to one or more of the PERCs based on the RVC codes listed for that visit. Because NHAMCS allows up to three complaints, each complaint was assigned to one of the PERCs. Clusters were combined, when appropriate, to ensure adequate numbers in each cluster for analysis (minimum cluster size=25). Those visits assigned to the PERC "Other" were reviewed to see if they could be assigned to a different existing PERC or if additional clusters would be warranted.

Because the reason for the visit is being used to predict the probability of requiring different levels of emergency care, the most useful classification would group complaints according to their likelihood of requiring care. The outcome of each visit in the NHAMCS 1998 and 2000 datasets was determined as described in section 4.C.ii, below. For each of the PERCs, the percentage of patients with each of the three outcomes was calculated. Each PERC was then assigned an ordinal ranking from 1 to 5 for each of the three outcomes, based on the quintile of risk for that outcome. For example, of all patients with the complaint of "fever," 39.3% received routine care, 56.5% received ED treatment, and 4.2% were admitted. These were in the 4th, 2nd, and 2nd quintiles of all complaint clusters for these three outcomes, respectively.

The existence of certain chronic medical conditions may increase the likelihood of a child requiring interventions in the ED. Past medical history was coded as a dichotomous variable indicating the presence or absence of medical conditions felt to be associated with need for emergency services, as determined a priori by the investigators. These conditions are:

Included conditions		**Common excluded conditions**
asthma includes: RAD/reactive airways disease, wheezing excludes: bronchiolitis	neuromuscular disease (muscular dystrophy, spinal muscular atrophy, myopathies)	ADD/ADHD
cardiac disease includes: any congenital heart disease, arrhythmia, or presence of a pacemaker excludes: mitral valve prolapse, "heart murmur"	renal failure/dialysis	constipation
cyclic vomiting	seizures/seizure disorder excludes: febrile seizures	developmental delay
cystic fibrosis	sickle cell disease	failure to thrive
diabetes mellitus	spinal cord injuries	
hemophilia or other bleeding disorder	tracheostomy	
inflammatory bowel disease (IBD)	transplant	
metabolic diseases (including congenital adrenal hyperplasia/CAH, urea cycle defects, galactosemia, organic acidurias)	VP shunt	

Similarly, current use of any prescription medications by patients, as recorded by the triage nurse, was considered. Use of prescription medications may reflect the existence of important medical conditions or may be an indication of prior treatment for the presenting problem. In either case, current medication use may influence subsequent care received in the ED.

Heart rate, respiratory rate, and blood pressure were adjusted for age by the use of age-appropriate normal values, obtained from the literature, and presented as a percentage of the age appropriate norm. Other objective information is often selectively recorded. For example, pulse oximetry is seldom measured at triage in children who do not have a complaint of a respiratory nature, as it is not clinically relevant. However, in the presence of respiratory complaints, pulse oximetry is likely to be clinically useful and an important predictor.[49,50] Performing pulse oximetry on all patients routinely would be extremely inefficient in terms of clinical care, yet excluding pulse oximetry as a predictor because it is selectively measured would lead to the loss of important information. Similarly, the Glasgow Coma Scale (GCS) is likely to be recorded only in those with head trauma or neurologic complaints, in whom it would be most relevant. We considered pulse oximetry and GCS to be normal, or uninformative, in those without an indication for their measurement. Indications for measurement were determined by the expert panel (see table below). For subjects with an indication but in whom pulse oximetry or GCS was not measured, the data were treated as missing. This follows the approach taken by the developers of the APACHE[32] and PRISM[14,51] scores for predicting ICU mortality, for which the assumption was made that variables that need to be measured will be

measured as indicated by patient care needs and that variables need not and should not be measured simply for prediction purposes.

chief complaints for which GCS is indicated	chief complaints for which pulse oximetry is indicated
altered mental status	any respiratory complaint in a patient with a past history of asthma
head injury	asthma/asthma attack
head trauma	breathing fast
ingestion/intoxication	difficulty or trouble breathing
	respiratory distress
	wheezing

Certain variables related to the timing of the visit may be related to the likelihood of interventions. For example, we have found, in a study of asthma, that the hospitalization rate for children with acute asthma is 60% higher during late night hours, despite 24-hour attending coverage in the ED; a patient survey found that those who use the ED at night have higher levels of concern about their child's illness and are somewhat more likely to be triaged as emergent.[52] Similarly, the use of diagnostic and therapeutic procedures might be modified by the seasonal prevalence of certain conditions such as influenza and RSV. Therefore, both time of day and season were examined as potential predictors of need for care. Time of day was categorized into three shifts, based on time of arrival in the ED: day (0800-1559), evening (1600-2359), and night (0000-0759). Season was categorized as winter, spring, summer, and fall based on the date of visit.

ii. Outcome variables. The outcome variable of interest is the level of care provided during the ED visit. The outcome variable is classified into three levels, as outlined below in Table 2. These levels were chosen to reflect an increasing need for care: routine nursing and physician assessment (including noninvasive monitoring and use of nonprescription medications); having diagnostic or therapeutic procedures performed in the ED but leading to discharge to home; and need for more intensive interventions requiring admission to the hospital or transfer to another facility. Death in the ED is combined with hospitalization in the most severe outcome category, because patients who die, if they had survived the ED visit, would undoubtedly be admitted or transferred; mortality is quite rare among pediatric emergency patients.

Table 2. Outcome Variables

Outcome level	Definition
Routine nursing and medical care	Discharged to home from ED; no diagnostic tests or therapeutic procedures performed
ED treatment	Discharged to home from ED; one or more of the following performed: diagnostic studies: • imaging studies (X-rays, ultrasound, CT, MRI) • lab tests on body fluids (including blood tests, urine tests, lumbar puncture) -- exclude urine pregnancy test or throat culture/rapid strep antigen test • tests not on body fluids (e.g., EKG, slit lamp exam) therapeutic procedures: • intravenous fluids • oxygen • prescription medications administered in ED (oral, IV, or inhaled) • wound management (suture repair, Steri-strip placement, burn dressing) • treatment of an orthopedic problem by splinting or casting, knee immobilizer, or crutches

	• specialty consultation • invasive diagnostic procedures (e.g., arthrocentesis, thoracentesis) resuscitation (CPR, bag-valve-mask ventilation)
Hospitalized	Admitted to hospital, transferred to another facility, or died in the ED

The initial plan was to analyze outcome as a single, three-category ordinal variable. However, subsequent analysis showed that the assumptions underlying the necessary ordinal model were not valid. We therefore chose to define two distinct outcome categories for purposes of analysis. One relevant grouping is to combine the categories of ED treatment and hospital admission, to predict the need for any care beyond routine nursing and physician care. Another clinically relevant analysis would be to examine the ability to predict need for hospitalization versus ability to be discharged from the ED (combining all other outcome categories). The two resulting outcome categories were:

a. any care: this included all patients who had ED treatment as defined above, PLUS all patients admitted to the hospital (i.e., all patients not categorized as routine)

b. admission: as defined in the table above

c. Cost and quality markers. A major goal of this study is to develop an acuity marker that can be used to adjust for risk in comparisons of other outcomes of care. One such outcome of interest is cost of care. Data on ED charges were obtained for a subset of patients in the study from their hospital bill. These charges were divided by each hospital's charge-to-cost ratio to yield an ED cost. Another relevant quality measure is ED length of stay, or throughput time: the time required to move a patient through the ED. Time from patient arrival until discharge was classified as total time, and time from being seen by a physician to leaving the department was classified as treatment time.

D. Data Collection Procedures and Logistics

All data on predictor and outcome variables were obtained from abstraction of the emergency department records. At each site, abstractors were hired and trained to review and abstract ED records. Scannable data forms were completed and mailed to the data management center for data entry. Range checks were incorporated into the data entry program to minimize entry errors, and all scanned forms were manually reviewed for accuracy. A 5% random sample of charts was reabstracted.

E. Data analysis.

1. Sample splitting. The total study sample was randomly divided into a derivation set (75% of the records) and a validation set (25% of the records). All the subsequently described analyses were performed first on the derivation set and then on the validation set, unless otherwise specified.

2. Weighting. Because of the complex sampling scheme, with different sampling fractions at each site, appropriate survey weighting methods were used. Analyses were performed using Stata version 8.0 (Stata Corp., College Station, TX). Probability weights were used, with clustering by site, to yield robust estimates of the standard errors.

3. Descriptive and univariate analyses. Predictor and outcome variables were described. Mean, standard deviation, median, range, and interquartile range were calculated for continuous variables.

For each potential predictor variable, the univariate association with the outcomes was determined. For each of the two dichotomous outcomes, the odds ratio for each predictor was calculated, with 95% confidence intervals. Ordinal variables were analyzed using chi-squared tests for trend and overall chi-squared tests. Plots were generated of each level of the ordinal variable versus the logit of the outcome. If the association appeared linear in the logit scale, the

ordinal variable was included in subsequent models as a single (linear) term; if there was evidence of nonlinearity, indicator variables were used in logistic models to represent different levels of the ordinal variable.

4. Multivariate analyses. Analyses in sections a. and b. to develop the multivariable models used the derivation set only. The model evaluations described in section c. were performed sequentially on both derivation and validation sets.

a. Selection of variables. Multivariable modeling techniques were used to develop predictive models for the outcomes of interest.[53,54] Although univariate screening is often used to reduce the number of variables to be included in a model, this practice is controversial. We have chosen to limit the number of candidate predictors *a priori* and to include all of the potential predictors in the initial model.

Although data were reasonably complete for most of the variables, the cumulative effect of missing data may be substantial if the data are missing more or less at random, leading to loss of important information.[54] If any given variable was missing in more than 15% of records, this variable was excluded from the modeling process. Two variables were included in the model as part of an interaction term that includes an indicator of whether the measurement of the variable was indicated: pulse oximetry (relevant only for subjects with respiratory illness) and Glasgow Coma Scale (relevant only in those with head trauma or a neurologic complaint).

Chief complaint was included in the model as follows: For each complaint listed for a given visit, the complaint was assigned the quintile ranking for each of the two outcomes (based on an independent dataset, the NHAMCS), as shown in Table 4. For example, a child with a complaint of head injury would have a value of 3 for that complaint in the model for receipt of care, and a 2 for admission For patients with more than one complaint, each complaint was assigned the appropriate ranking, and the highest value for all the complaints for a given patient was entered into the model. Thus, a child with complaints of fever and earache would be assigned rankings of 2 and 1, respectively, for need for admission; the value of 2 would be used for that patient in the model for admission.

b. Modeling strategies. The outcome, as shown in Table 2 above, has three levels representing mutually exclusive and collectively exhaustive categories with an implicit ordering, indicative of the intensity of services provided and of the underlying severity of illness.

For polytomous outcomes, a generalization of the usual logistic regression modeling that handles ordinal responses can be used.[55,56] However, we found that our multicategory outcome does not have an ordered relationship with the predictor variables of interest (specifically, the assumption of proportional odds was violated),[55] making this modeling technique inappropriate for the data. We therefore employed standard logistic regression to estimate two separate models, one for each of the outcomes of any care and admission.

For all logistic regression analyses, parameters were estimated using maximum likelihood techniques. First the full model was fitted. Predictor variables with a nonsignificant ($p>0.10$) association with the outcome category were dropped sequentially from the model, and the resulting model was compared with the more complete one. The Bayesian information criterion was calculated for each model, and a difference of less than 6 was indicative of insufficient evidence to reject the more parsimonious model.[55]

c. Model evaluation. A challenge of evaluating the results of regression models using a complex sampling scheme is the lack of readily available methods for determining model fit. A reasonable alternative, which we employed, is to estimate model parameters taking the survey design into account (“design-based” method) but to use a model-based (i.e., no weights) methods for determining model fit, as suggested by Hosmer and Lemeshow.[56, pp.219-220]

For each model, the adjusted r-squared was calculated. Calibration of the models was determined by the Hosmer-Lemeshow goodness of fit test. For this test, a p value less than 0.05 was considered evidence of lack of fit. Discriminative ability of the models was assessed using the c statistic, or area under the receiver-operator characteristic curve.

d. Sample size. The projected sample size was calculated based on the ability to develop a model for the least common outcome – admission. We wished to have adequate sample size for 1) inclusion of the potential predictor variables in the mode; 2) estimation of model sensitivity and specificity within a reasonable degree of precision; and 3) detection of moderate association between uncommon predictors and the outcome.

Using the general rule of thumb of needing at least 10 outcomes for each degree of freedom included in a logistic regression model, we would need at least 210 subjects admitted in the derivation subsample to accommodate all of the potential predictors. Assuming an overall admission rate of 7%, and 80% of subjects with complete data on the predictors, we would need 3750 subjects in the derivation sample, or a total of 5000 subjects. This would yield 100 admitted subjects in the validation sample. With this sample size, sensitivity for predicting admission would be estimated with a maximum 95% confidence interval width of ± 0.1. Finally, this proposed sample size would allow detection of a relative risk for a moderately common predictor of approximately 1.5; for predictors with a low prevalence (10%), the detectable RR is 1.8. Therefore, a target of 5000 total subjects was desired. To ensure adequate representation of subjects from all sites, a target of 1000 subjects was set for the two general hospitals and 1500 from each of the pediatric hospitals.

F. Limitations.

Any retrospective study is limited by the availability and quality of data recorded in the medical record. Important predictors may not be recorded routinely; missing data may lead to bias, or loss of relevant predictive information. Because the RePEAT would most likely be applied retrospectively to existing data sources, however, the study procedure is relevant to actual intended practice.

Our study was conducted at only four institutions. If the practice at these hospitals is not representative of generally accepted standards, our results may not be generalizable. External validation in other settings will be necessary.

Our hypothesis is that level of care in the ED can be predicted from a relatively small subset of clinical information available at the time of triage. The predictive ability of such information is likely to be modest. However, we believe this is offset by the advantage of being able to apply the score to readily available, easily obtained data.

5. RESULTS.

A. Study Population

A total of 5521 subjects was enrolled (Table 3). The sample size was approximately 10% higher than originally estimated, because ED volume at all hospitals increased from the previous year, on the basis of which sampling fractions were determined. The dataset was split into a derivation set (n=4421) and a validation set (n=1361).

Table 3. Study Population

			Race/ethnicity			Insurance	
site	n	mean age (yrs)	% non-Hisp. white	% black	% Hispanic	% public insurance	% self pay
101	1851	5.5	41.7%	39.3%	12.6%	36.1%	7.2%
201	1174	8.6	27.5%	53.0%	<1%	55.0%	10.4%
301	1521	5.6	55.4%	26.0%	5.5%	37.2%	5.8%
401	975	7.7	78.0%	13.6%	6.2%	42.7%	7.5%
OVERALL	5521						

B. PERCs

A total of 53 PERCs was created. The clusters are shown in Table 4, along with their quintile ranking of likelihood of receiving any care and likelihood of admission (based on the 1998 and 2000 NHAMCS data).

Table 4. Pediatric Reason for Visit Clusters

Reason for Visit Cluster	Quintile for receiving care	Quintile for admission	Reason for Visit Cluster	Quintile for receiving care	Quintile for admission
Abdominal pain	4	4	Follow-up/recheck	3	4
Abuse/assault	3	3	General/unspec. sx	3	4
Allergic reaction	2	3	GI bleeding	5	3
Alt. mental status	5	5	Gynecologic	4	2
Asthma/wheezing	4	4	Head/neck trauma	3	2
Bites/stings	4	1	Headache	3	2
Burn	5	3	Laceration	5	1
Cardiac complaint	5	5	Lump/mass	2	1
Chest pain	4	3	Male GU	4	2
Chronic disease	5	5	Multiple trauma	4	3
Congestion/URI	1	2	MVC	3	3
Constipation	1	1	Neck pain	3	2
Cough	2	2	Neuro (other)	4	4
Croup	4	1	Other	3	4
Crying/colic	1	2	Poisoning	2	4
Dental complaint	3	2	Poor feeding/mouth complaints	2	2
Device complication	3	4	Pregnancy	4	5
Diarrhea	2	3	Primary care	3	3
Ear complaints	1	1	Psych/behavioral	2	5
Epistaxis	2	2	Rash	2	2
Extremity pain or injury	5	2	Respiratory (other)	4	4
Eye complaints	1	2	Seizure	4	4
Fainting/syncope	4	3	Sore throat	3	2
FB (ENT/GI)	1	3	Trauma – other/unsp.	4	2
FB (skin)	4	2	Urinary symptoms	4	2
Fever	2	2	Vomiting	3	3
Fever – neonate	3	5			

C. Outcomes.

Table 5 shows the distribution of outcomes in the derivation and validation samples.

Table 5. Outcomes in the study sample

	Derivation		Validation	
Site	% any care	% admitted	% any care	% admitted
101	61.8	11.1	61.3	11.4
201	68.2	1.9	63.6	3.1
301	71.1	11.5	71.3	11.1
401	70.8	3.3	73.7	2.5
Overall	67.3	7.9	66.5	7.9

D. Predictors and univariate measures of association.

The values for predictor variables are summarized in Table 6 below along with the univariate measure of association for each predictor and the two outcome variables. Several potential predictors were dropped from further consideration due to missing data. Blood pressure was documented only 45% of the time overall, including 66% of the time for children

3 years and older and 20% for those younger than 3. Glasgow Coma Score (GCS) was documented in only 55% of cases in which measurement of GCS would have been indicated. In addition, even when indicated, GCS was rarely abnormal (5% of patients). These two variables, therefore, were eliminated from the modeling process. Conversely, pulse oximetry data was present in 89% of cases meeting the definition for having pulse oximetry indicated.

Table 6. Univariate associations

Predictor variable	Mean ± SD or proportion	Univariate OR for care received (95% CI)	Univariate OR for admission (95% CI)
Age	6.7 ± 5.7 yrs		
<3 mos.	4.6%	reference group	reference group
3-24 mos.	24.6%	0.87 (0.73, 1.03)	0.41 (0.25, 0.70)
>24 mos.	70.8%	1.59 (1.47, 1.72)	0.31 (0.23, 0.40)
Past medical history	32.4%	1.91 (1.58, 2.31)	2.28 (1.89, 2.75)
Current medications	31.5%	1.67 (1.30, 2.14)	2.10 (1.67, 2.63)
Mode of transportation	6.9%	3.59 (1.82, 7.07)	5.00 (3.28, 7.60)
Triage status			
nonurgent	55.4%	reference group	reference group
urgent	39.1%	3.08 (1.45, 6.56)	8.59 (1.92, 38.3)
emergent	5.5%	17.64 (4.70, 66.11)	39.50 (5.66, 275.54)
Heart rate (as multiple of age-appropriate norm)	1.06 ± 0.21	2.90 (0.74, 11.4)	12.86 (6.16, 26.81)
Respiratory rate (as multiple of age-appropriate norm)	1.11 ± 0.34	1.83 (1.37, 2.44)	3.91 (2.34, 6.56)
Temperature (ºC)	37.2 ± 1.1	1.00 (0.96, 1.05)	1.21 (1.15, 1.27)
Oxygen saturation	96.6 ± 3.5	1.01 (1.008, 1.012)	1.013 (1.01, 1.15)
Season			
winter	28.1%	reference category	reference category
spring	22.6%	1.15 (0.91, 1.46)	1.34 (0.75, 2.39)
summer	26.5%	1.02 (0.85, 1.22)	1.36 (0.58, 3.19)
fall	22.8%	1.10 (0.88, 1.36)	1.38 (0.70, 2.69)
Time of day			
night	11.3%	reference category	reference category
day	35.5%	1.07 (0.94, 1.23)	0.93 (0.79, 1.10)
evening	53.2%	0.84 (0.62, 1.13)	0.91 (0.74, 1.11)

E. Model development

We first examined the possibility of developing a single model with an ordinal outcome of levels of care using the *ologit* procedure. We used the approximate likelihood-ratio test of proportionality of odds across response categories, implemented as the *omodel* command in Stata. The resulting p value was <0.0001, indicating strong evidence against the proportional odds assumption. We therefore proceeded with standard logistic regression, as outlined in section 4.E.4.b above. We developed two models, one to predict receipt of any ED care and the other to predict admission.

The modeling process began with all predictor variables in Table 6 plus the PERC ranking. As outlined above, nonsignificant predictors were removed sequentially, and the Bayesian information criterion was calculated. A difference of <6 in the BIC for two models resulted in keeping the variable out of the model.

In total, 3427 subjects (83%) in the derivation set had complete data on all predictors and outcomes and were included in the modeling process. The results of the two models are shown in the tables below.

Table 7. Logistic regression model for receipt of ED care

variable	logit coefficient	OR	95% CI for OR	p value
complaint ranking				
1		ref.		
2	0.064	1.07	0.91, 1.24	0.411
3	0.400	1.49	1.28, 1.74	<0.001
4	1.172	3.23	2.49, 4.18	<0.001
5	2.006	7.43	5.81, 9.51	<0.001
age group				
<3 mos.		ref		
3-36 mos.	-0.103	0.90	0.85, 0.95	<0.001
>36 mos.	0.416	1.52	1.27, 1.81	<0.001
triage category				
nonurgent		ref		
urgent	0.868	2.38	1.90, 2.98	<0.001
emergent	2.326	10.23	5.29, 19.78	<0.001
Arrival via EMS	0.686	1.99	1.43, 2.75	<0.001
Currently taking prescription medications	0.332	1.39	1.09, 1.78	0.007
HR (as multiple of age-appropriate norm)	0.667	1.95	0.86, 4.41	0.11
RR (as multiple of age-appropriate norm)	0.376	1.46	1.18, 1.80	0.01
Temperature (°C)	0.172	1.19	1.11, 1.27	<0.001
Constant	-8.201	---	---	<0.001

model pseudo-R^2 = 0.1524

Table 8. Logistic regression model for admission

variable	logit coefficient	OR	95% CI for OR	p value
complaint ranking				
1		ref.		
2	1.672	5.32	0.49, 57.98	0.17
3	1.696	5.44	0.56, 52.86	0.144
4	2.294	9.92	1.13, 86.78	0.038
5	2.735	15.40	1.54, 153.41	0.02
age group				
<3 mos.		ref		
3-36 mos.	-0.681	0.51	0.33, 0.77	0.002
>36 mos.	-1.134	0.32	0.19, 0.56	<0.001
triage category				
nonurgent		ref		
urgent	1.784	5.95	2.17, 16.33	0.001
emergent	3.146	23.23	6.31, 85.50	<0.001
Arrival via EMS	0.649	1.91	1.74, 2.10	<0.001
Currently taking prescription medications	0.342	1.41	1.23, 1.60	<0.001
HR (as multiple of age-appropriate norm)	1.233	3.43	1.14, 10.31	0.028
RR (as multiple of age-appropriate norm)	0.408	1.50	1.11, 2.03	0.007
Constant	-6.738	---	---	0.001

model pseudo-R^2 = 0.2660

F. Model assessment

The goodness of fit of each model was determined in both the derivation and validation samples. As shown in Table 9, the fit of both models was acceptable in both datasets. It should be noted that the goodness-of-fit test was performed without the probability weights; hence, the total number of observed and expected outcomes may not match closely.

Table 9. Goodness of fit.

	Receipt of care model				Admission model			
	Derivation		Validation		Derivation		Validation	
quintile	observed	expected	observed	expected	observed	expected	observed	expected
1	112	106	45	37.7	3	1.5	1	0.5
2	141	133.2	46	47.8	4	3.3	1	1.2
3	172	155.2	62	56.9	5	4.4	2	1.5
4	179	179.3	61	64.6	7	5.6	1	2.0
5	197	204.9	66	72.8	3	7.9	1	2.6
6	241	226.6	86	79.8	10	14.4	4	4.6
7	244	246.8	84	88.1	17	24.0	6	8.0
8	267	265.7	97	94.2	32	37.0	11	13.3
9	275	283.2	99	101.3	53	61.1	17	22.7
10	294	301.7	107	107.8	133	136.9	43	48.2
TOTAL	2122	2102.6	753	751	267	296.1	87	104.6

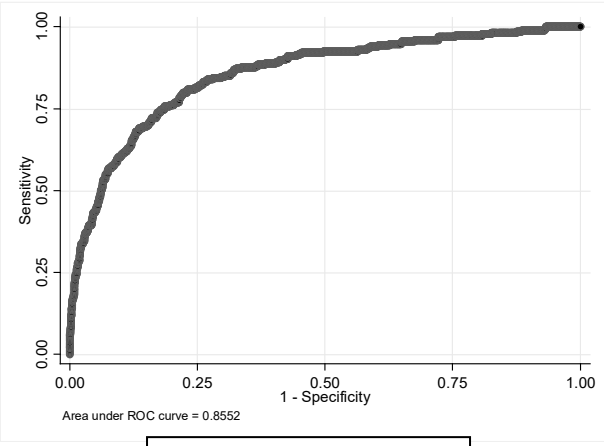
Hosmer-Lemeshow goodness of fit chi-squared (derivation): 14.6, 8 d.f., p=0.09 for receipt of care model (validation): 8.74, 10 d.f., p=0.68

Hosmer-Lemeshow goodness of fit chi-squared (derivation): 11.2, 8 d.f., p=0.27
for admission model (validation): 6.00, 10 d.f., p=0.88

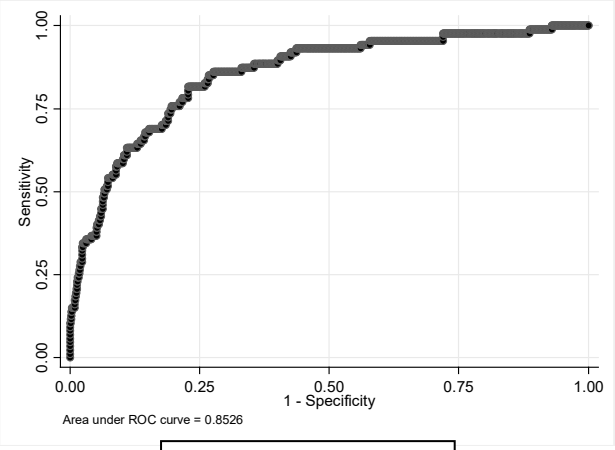
We also sought to evaluate the ability of the models to discriminate those with and without the outcome of interest. The resulting ROC curves are shown below. For admission, the discrimination is excellent, and for receipt of care, it is good, as shown by the c statistics. For both models, the discriminative performance was similar in both the validation and derivation subsets.

Figure 1. ROC curves

a. admission

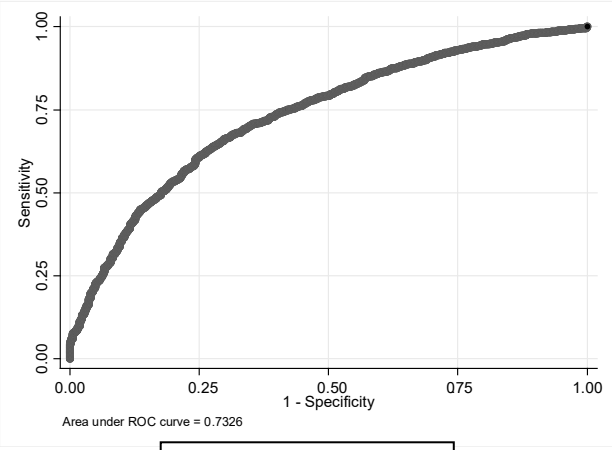


c statistic = 0.8552

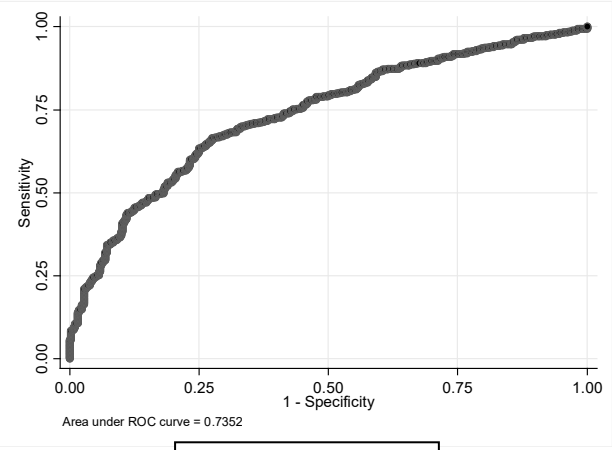


c statistic = 0.8526

b. receipt of care



c statistic = 0.7326



c statistic = 0.7352

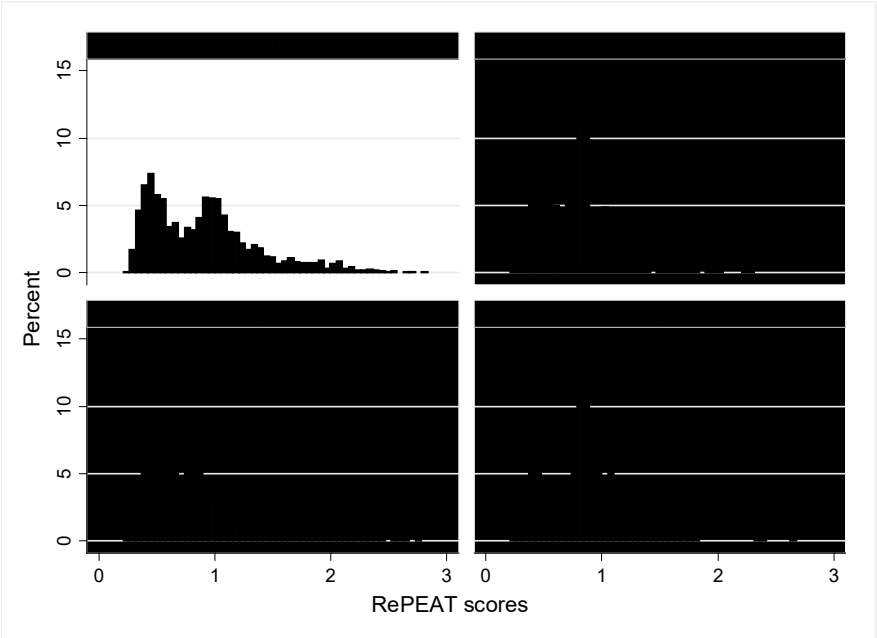
G. RePEAT scores

The RePEAT score was defined as the sum of the predicted probability of receiving care plus twice the predicted probability of admission. The overall distribution of RePEAT scores is shown in Table 10, and the distribution at each site is shown in Figure 2.

Table 10. Distribution of RePEAT scores

mean ± SD	0.830 ± 0.394
median [interquartile range]	0.792 [0.522, 1.013]
total range	0.213, 2.834

Figure 2. Distribution of RePEAT scores by site



As shown in the figure, there is a good spread of scores at each site. Sites 101 and 301 have a flatter distribution, reflecting a greater number of patients with higher predicted resource use.

H. RePEAT scores and other outcomes

As noted above, one potential use of the RePEAT score is to adjust for differences in severity when comparing outcomes across settings. We examined two outcomes for which severity adjustment may be useful: charges and throughput time. Three sites provided charge data on a subset of their patients; at one site, concerns that linkage of charge data would violate HIPAA (which was enacted during the study) prevented obtaining financial data.

Table 11 shows the results of linear regressions. In all cases, the association between RePEAT score was statistically significant and potentially clinically meaningful. The smallest association was between total throughput time (i.e., arrival to discharge) and RePEAT score. The correlation coefficient was 0.25, indicating a relatively modest association; RePEAT score explained only 6% of the variation in length of stay. However, for the other outcomes, RePEAT score explained 13 to 16.5% of the variation.

Table 11. Association of RePEAT with financial and time outcomes

variable	regression coefficient (95% CI)	p value	model R^2
hospital charges	449.9 (363.3, 536.5)	<0.001	0.1654
hospital costs	156.9 (126.5, 187.2)	<0.001	0.1609
total time	0.94 (0.80, 1.08)	<0.001	0.0598
treatment time	1.26 (1.10, 1.41)	<0.001	0.1305

We examined the potential usefulness of risk adjusting the comparisons of these outcomes across sites using the RePEAT score. For each outcome, the difference between sites was modeled using linear regression with a single dummy variable for site, and then a term for RePEAT score was added. In all comparisons, the explanatory power of the model was improved with the addition of the severity score. In addition, the magnitude of the differences is substantially different when adjusted for severity. Site 201, for example, has markedly lower charges than the reference site 101 (which is in the same geographic region). However, the difference is much smaller when adjusted for severity. Similarly, sites 401 and 201 appear to be much more efficient in moving patients through the ED compared with site 101, but the risk-adjusted difference shows the sites to be more comparable. If we examine site 301 vs.101, a different pattern emerges. Both are similar settings, but the number of beds per patient is smaller at site 301, which therefore would be expected to be more crowded. The total throughput time is indeed higher at site 301, but the risk-adjusted difference is even greater than the raw numbers suggest.

Table 12. Risk-adjusted comparisons of financial and time outcomes across sites

	Hospital charges		Hospital costs		Total time		Treatment time	
site (site 101 is reference category)	raw difference	diff. adjusted for RePEAT	raw difference	diff. adjusted for RePEAT	raw difference	diff. adjusted for RePEAT	raw difference	diff. adjusted for RePEAT
201	-159.18	-86.78	-41.46	-16.48	-0.75	-0.51	---**	---
301	---*	---	---	---	0.93	1.09	0.55	0.81
401	91.26	148.69	64.83	83.10	-0.79	-0.65	-0.35	-0.11
model R^2	0.0087	0.1706	0.0090	.1683	0.1113	0.1843	0.0382	0.1987

* financial data not provided from site 301

** MD time not recorded at site 201, so treatment time could not be calculated

I. Discussion and Significance.

We have shown that a model based on a small number of variables, routinely obtained at the time of triage, accurately predicts the level of resource utilization in the emergency department for pediatric patients. The predicted probabilities from this model can be used as a marker of severity; patients with higher predicted resource utilization presumably are more severely ill. Moreover, the predicted probabilities can be used to adjust for differences in baseline risk when comparing other outcomes and quality markers, such as costs and length of stay, across settings.

Advantages of the RePEAT score include parsimony (a total of eight variables) and near universality of data availability (89% of charts abstracted at the four sites contained complete information on all predictors in the model). This makes the RePEAT potentially applicable in a wide variety of settings and may be amenable to application to large electronic datasets.

An alternative risk-adjustment tool for pediatric emergency visits has recently been developed and validated. The PRISA uses a similar approach in that it predicts probability of admission. The PRISA score includes 18 variables, including three laboratory variables and three treatment variables. We believe that the RePEAT has some advantages over the PRISA. First, the laboratory variables included in the PRISA are obtained in only a very small minority of all ED patients. The PRISA assumes an unmeasured variable to be normal, which may lead to substantial bias.[57] Moreover, performance of laboratory tests and therapeutic interventions determine, in part, the PRISA score; using such a score to adjust for differences in costs and time variables (both of which are related to the performance of such procedures) is not appropriate. Finally, admission is a relatively uncommon outcome. A score based only on admission risk may not discriminate well among the large proportion of patients at low risk.

However, there are likely situations in which the more detailed clinical information contained in the PRISA may be preferable. Among a subset of children at reasonably high risk of admission, the basic triage information included in RePEAT may be insufficient to discriminate various levels of risk. Thus, the scores may provide complementary information in certain situations, such as subsets of patients with high-acuity diagnoses or in higher-acuity settings. The relative usefulness of these two risk-adjustment tools for different purposes would be a fruitful area for future study.

J. Limitations

Several limitations have already been mentioned. In addition, it must be noted that the regression coefficients were determined in a small sample of EDs. Although there is substantial diversity in patient populations and settings, the sample of EDs in this study is not necessarily representative. This is somewhat mitigated by the use of an external, nationally representative dataset (NHAMCS) to derive quantiles of risk for the chief complaints. However, further external validation of the RePEAT score is necessary before widespread adoption can be recommended.

K. Implications and future directions.

The RePEAT is a potentially valuable tool for risk adjustment in studies of outcomes and quality of pediatric emergency care. We are currently beginning a study to further validate and refine the tool in a larger, more representative dataset. We then plan to utilize this tool in evaluating quality of pediatric emergency care.

6. PUBLICATIONS AND PRODUCTS

A. Publications

Gorelick MH, Lee C, Cronan K, Kost K, Palmer K. Pediatric emergency assessment tool (PEAT): a risk-adjustment measure for pediatric emergency patients. *Acad Emerg Med* 2001;8:156-162.

Gorelick MH. Severity of illness measures for pediatric emergency care: are we there yet? [editorial]. *Ann Emerg Med* 2003;41:639-643.

Gorelick MH. Bias arising from missing data in predictive models. [manuscript submitted]

B. Presentations

Gorelick MH, Cronan K, Kost S, Palmer K, Lee C. PEAT: Pediatric Emergency Assessment Tool. (Presented at the Annual Meeting of the Ambulatory Pediatric Association, Boston, MA, May 2000, and the Annual Meeting of the Society for Academic Emergency Medicine, San Francisco, May 2000

Gorelick MH. Bias arising from missing data in predictive models. *Acad Emerg Med* 2002; 9:483-484. (Presented at the Annual Meeting of the Pediatric Academic Societies, Baltimore, MD, May 2002, and the Annual Meeting of the Society for Academic Emergency Medicine, St. Louis, MO, May 2002)

Gorelick MH, Alpern ER, Alessandrini EA. A system for grouping pediatric emergency department complaints, the Pediatric Emergency Reason for Visit Clusters (PERC). Pediatr Res 2004;55:119A. (Presented at the Annual Meeting of the Pediatric Academic Societies, San Francisco, CA, May 2004.)

7. REFERENCES

[1] Seidel JS, Henderson DP, eds. *Emergency medical services for children: a report to the nation*. Washington DC: National Center for Education in Maternal and Child Health, 1991.

[2] Jaffe DM. The Institute of Medicine report on emergency medical services for children: research in emergency medical services for children. *Pediatrics* 1995;96(1S):191-194.

[3] Durch JS, Lohr KN, eds. *Emergency medical services for children. Institute of Medicine report.* Washington DC: National Academy Press, 1993.

[4] Young WW, Swinkola RB, Zorn DM. The measurement of hospital case-mix. *Med Care* 1982;20:501-512.

[5] Vladeck BC, Kramer PS. Case mix measures: DRGs and alternatives. *Ann Rev Pub Health* 1988;9:333-359.

[6] Gross PA, Beyt BE, Decker MD et al. Description of case-mix adusters by the severity of illness working group of the Society of Hospital Epidemiologists of America (SHEA). *Infect Control Hosp Epidemiol* 1988;9:309-316.

[7] Cales RH. Injury severity determination: requirements, approaches, and applications. *Ann Emerg Med* 1986;15:1427-1433.

[8] Ruttiman UE, Pollack MM. Variability in duration of stay in pediatric intensive care units: a multiinstitutional study. *J Pediatr* 1996;128:35-44.

[9] Gorelick MH. Severity of illness measures for pediatric emergency care: are we there yet? [editorial]. *Ann Emerg Med* 2003;41:639-643.

[10] Gibson G. Indices of severity for emergency medical evaluative studies: reliability, validity, and data requirements. *Int J Health Services* 1981;11:597-630.

[11] Blumberg MS. Risk adjusting health care outcomes: a methodologic review. *Med Care Rev* 1986;43:351-393.

[12] Korin JB, Slebst SM. Legal aspects of emergency department pediatrics. In Fleisher GR, Ludwig S, eds. *Textbook of pediatric emergency medicine*, 3rd. ed. Baltimore: Williams and Wilkins, 1993; 1544-64.

[13] Pollack MM, Patel KM, Ruttiman UE. PRISM III: an updated pediatric risk of mortality score. *Crit Care Med* 1996;24:743-752.

[14] Pollack MM, Ruttiman UE, Getson PR. Pediatric risk of mortality (PRISM) score. *Crit Care Med* 1988;16:1110-6.

[15] Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE -- acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 1981;9:591-7.

[16] Lemeshow S, Teres D, Pastides H, Avrunin JS, Steingrub JS. A method for predicting survival and mortality of ICU patients using objectively derived weights. *Crit Care Med* 1985;13;519-525.

[17] Baker SP, O'Neill B, Haddon W et al. The Injury Severity Score: a method for describing patinets with multiple injuries and evaluating emergency care. *J Trauma* 1974;14:187-196.

[18] Brewster AC, Karlin BG, Hyde LA, Jacobs CM, Bradbury RC et al. MEDISGRPS™: a clinically based approach to classifying hospital patients at admission. *Inquiry* 1985;22:377-87.

[19] Gustafson DH, Fryback DG, Rose JH, Prokop CT, Detmer DE, Rossmeissl JC et al. An evaluationof multiple trauma severity indices created by different index development strategies. *Med Care* 1983;21:674-691.

[20] Rutledge RR. The Injury Severity Score is unable to differentiate between poor care and severe injury. *J Trauma* 1996;944-950.

[21] Maslanka AM. Scoring systems and triage from the field. *Emerg Med Clin North Amer* 1993;11:15-27.

[22] Orr RA, Venkataraman ST, McCloskey KA, King WD, Cinoman MA. Predicting the need for major interventions during pediatric interhospital transport using pre-transport variables [abstract] *Pediatr Emerg Care* 1992;8:371.

[23] Koehler JJ, Baer LJ, Malafa SA et al. Prehospital index: a scoring system for field triage of trauma victims. *Ann Emerg Med* 1986;15:178-82.

[24] Brillman JC, Doezema D, Tandberg D et al. Triage: limitations in predicting need for emergency care and hospital admission. *Ann Emerg Med* 1996;27:493-500.

[25] Chamberlain JM, Patel KM, Ruttiman UE, Pollack MM. Pediatric risk of admission (PRISA): a measure of severity of illness for assessing the risk of hospitalization from the emergency department. *Ann Emerg Med* 1998;32:161-9.

[26] Cullen DJ, Civetta JM, Briggs BA, Ferrara LC. Therapeutic intervention scoring system: a method for quantitative comparison of patient care. *Crit Care Med* 1974;2:57-60.

[27] Leape LL, Park RE, Solomon DH, Chassin MR, Kosecoff J, Brook RH. Does inappropriate use explain small-area variations in the use of health care resources. *JAMA* 1990;263:669-672.

[28] Merritt TA, Palmer D, Bergman DA, Shiono PH. Clinical practice guidelines in pediatric and newborn medicine: implications for their use in practice. *Pediatrics* 1997;99:100-114.

[29] Yeh TS, Pollack MM, Holbrook PR, Fields AI. Assessment of pediatric intensive care -- application of the Therapeutic Intervention Scoring System. *Crit Care Med* 1982;10:497-500.

[30] Abbuhl SB, Lowe RA. The inappropriateness of "appropriateness" [editorial]. *Acad Emerg Med* 1996 3:189-191.

[31] Gill JM. Nonurgent use of the emergency department: appropriate or not? *Ann Emerg Med* 1994;24:953-957.

[32] Knaus WA, Wagner DP, Draper EA et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100:1619-36.

[33] Strauss RH, Rooney B. Critical care pediatrician-led aeromedical transports: physician interventions and predictiveness of outcome. *Pediatr Emerg Care* 1993;9:270-274.

[34] Orr RA, Venkataraman ST, Cinoman MI, Hogue BL, Singelton CA, McCloskey KA. Pretransport Pediatric Risk of Mortality (PRISM) score underestimates the requirement for intensive care of major interventions during interhospital transport. *Crit Care Med* 1994;22:101-107.

[35] MacKenzie EJ. Injury severity scales: overview and directions for future research. *Am J Emerg Med* 1984;2:537-549.

[36] van der Windt DAM, Nagelkerke AF, Bouter LM, Dankert-Roelse JE, Veerman AJ. Clinical scores for acute asthma in pre-school children: a review of the literature. *J Clin Epidemiol* 1994;47:635-646.

[37] Super DM, Cartelli NA, Brooks LJ, Lembo RM, Kumar ML. A prospective randomized double blind study to evaluate the effect of dexamethasone in acute laryngotracheobronchitis. *J Pediatr* 1989;115:323-9.

[38] McCarthy PL, Sharpe MR, Spisel SZ et al. Observation scales to identify serious illness in febrle children. *Pediatrics* 1982;70:802-9.

[39] Tritschler V. Emergency department nursing. In Wilkins WE, ed. *Emergency medicine: scientific foundations and current practice*, 3rd ed. Baltimore: Williams and Wilkins, 1989; 977-980.

[40] Wilson FP, Wilson LO, Butler AB, Canales L. Algorithm-directed triage of pediatric patients: a prospective study. *JAMA* 1980;243:1528-1531.

[41] Wilson LO, Wilson FP, Wheeler M. Computerized triage of pediatric patients: Automated triage algorithms. *Ann Emerg Med* 1981;10:636-640.

[42] Wilson FP, Wilson LO, Canales L. Algorithm-directed triage of pediatric patients: a retrospective study. *Ann Emerg Med* 1981;10:427-431.

[43] George S, Read S, Westlake L, Fraser-Moodie A, Pritty P, Williams B. Differences in priorities assigned to patients by triage nurses and by consultant physicians in accident and emergency departments. *J Epidemiol Comm Health* 1993;47:312-315.

[44] George S, Westlake L, Williams B, Fraser-Moodie A, Pritty P. Evaluation of nurse triage in a British accident and emegency department. *BMJ* 1992;304:876-8.

[45] Lowe RA, Bindman AB, Ulrich SK, Norman G, Scaletta TA, Keane D, Washington D, Grumbach K. Refusing care to emergency department patients: Evaluation of published triage guidelines. *Ann Emerg Med* 1994;23:286-293.

[46] Chamberlain JM, Patel KM, Pollack MM, Brayer A, Macias CG, Okada P, Schunk JE for the Collaborative Research Committee of the Emergency Medicine Section of the American Academy of Pediatrics. Recalibration of the Pediatric Risk of Admission score using a multi-institutional sample. *Ann Emerg Med* 2004;43:461-468.

[47] Reason for Visit Classification and Coding Manual, 1997. National Ambulatory Medical Care Survey. Hyattsville, Maryland: National Center for Health Statistics. 1997. (http://www.cdc.gov/nchs/data/ahcd/rvc97.pdf)

[48] Stussman BJ. National Hospital Ambulatory Medical Care Survey: 1994 Emergency Department Summary. Advance data from vital and health statistics; no.275. Hyattsville, MD: National Center for Health Statistics, 1996.

[49] Margolis PA, Ferkol TW, Maroscol S, Super DM, Keyes LL, McNutt R, Harrell FE. Accuracy of the clinical examination in detecting hypoxemia in infants with respiratory illness. *J Pediatr* 1994;124:552-560.

[50] Anderson AB. Zwerdling RG. Dewitt TG. The clinical utility of pulse oximetry in the pediatric emergency department setting. *Pediatr Emerg Care* 1991;7:263-6.

[51] Yeh TS, Polack MM, Ruttiman UE, Holbrook PR, Fields AI. Validation of a physiologic stability index for use in critically ill infants and children. *Pediatr Res* 1984;18:445-451.

[52] Pachter LM, Ludwig S, Groves A. Night people: utilization of a pediatric emergency department during the late night. *Pediatr Emerg Care* 1991;7:12-14.

[53] Ruttiman UE. Statistical approaches to development and validation of predictive instruments. *Crit Care Clinics* 1994;10:19-35.

[54] Harrell FE, Lee KL, Mark DB Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-387.

[55] Long JS. Regression models for categorical and limited dependent variables. Thousand Oaks, CA: SAGE Publications, 1997.

[56] Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: Wiley and Sons, Inc. 1989.

[57] Gorelick MH. Bias arising from missing data in predictive models [abstract]. *Acad Emerg Med* 2002; 9:483-484.