# Statistical Analysis Plan

## The Right Call:
## An Implementation Trial of a Toolkit to Improve Diagnosis of Sepsis during Pediatric Transfer Phone Calls

PI:
Halden F. Scott, MD, MSCS

Research Institutions:
University of Colorado Anschutz
Children's Hospital Colorado
Adult and Child Center for Outcomes Research & Delivery Science

ClinicalTrials.gov Number:
#NCT07051668

Protocol:
COMIRB #23-2340

Authors:
Carter Sevick, MS, PhD
Halden Scott, MD, MSCS

Version: 1
Date: July 7, 2025

## 1. Introduction

Pediatric sepsis is a leading cause of childhood mortality, in which first-hours, evidence-based emergency care can prevent death and improve outcome. In the US, >70,000 children are hospitalized with sepsis yearly, with a nationwide cost of $7.31 billion. Case fatality rates range from 5-10%; 35% of survivors experience diminished quality of life a year after sepsis, and increased healthcare utilization. Outcomes are improved by timely diagnosis and standardized first hours' resuscitative care; however this care is not always delivered.

Most children with sepsis are treated in general Emergency Departments (EDs), which have worse outcomes. Diagnosing sepsis is challenging everywhere, particularly in children in general EDs. Delays in diagnosing sepsis have been implicated as a driving factor in preventable sepsis mortality, including in children. Children in whom the diagnosis of sepsis is missed in the presenting ED are less likely to receive timely treatments including antibiotics and fluids, and are more likely to require inpatient escalations of care and have longer hospital lengths of stay. Sepsis is less likely to be diagnosed when present in general EDs, where providers and nurses see fewer children. General ED physicians and nurses have described the difficulty and variation in diagnosing sepsis based on an individual provider's pediatric experience and acumen.

Phone calls to accepting subspecialty pediatric ED physicians are legally mandated and universally occur before transfer of a patient out of a general ED in the United States but are under-studied. The Emergency Transfer and Active Labor Act (EMTALA) requires specialized pediatric hospitals, which legally represent a higher level of care for any child in a general ED, to provide subspecialty consultation to the general ED. To facilitate these conversations, children's hospitals have established 24-hour call centers that provide live consultation from pediatric emergency physicians to providers in general EDs. This represents an important moment for knowledge transfer. In multiple studies, general emergency physicians refer to the importance of the real time call to a pediatric emergency physician at a specialized children's hospital to receive live advice as an important element of their care for emergency care for children.

Our quantitative work identified lower rates of diagnosing children with sepsis when sepsis is present in general and community pediatric EDs compared to pediatric tertiary EDs. When the diagnosis of sepsis is not formally made (using the term in notes, or using a sepsis order set or pathway), our research has shown that sepsis treatments are more likely to be delayed, and outcomes are worse. We have explored the current state, barriers and facilitators to pediatric sepsis care in qualitative studies with physician and nursing leaders. They identified diagnosis as one of the biggest challenges they face. They also identified that children's hospitals played an important role in supporting (or hindering) their care, describing the transfer call in quotes that suggest some of the interpersonal dynamics between physicians that influence such calls, such as, "I think… if someone's calling you for help, they probably need help, and just be nice with it."

The study team, in collaboration with invested clinicians and expert partners, has developed a pediatric sepsis diagnostic safety toolkit that will be implemented in a pediatric health system's transfer call center. Preparation for launch of the toolkit will include education throughout the health system, with a focus on transfer center nurses and accepting emergency and critical care physicians who will be partnering in delivering the toolkit. Usual avenues for clinical education will be used, including meetings, endorsement from clinical

leaders, emails, and physical materials such as badge cards. Referring ED providers outside of the pediatric health system will receive materials containing evidence-based pediatric sepsis diagnostic criteria through existing pediatric preparedness outreach programs.

The toolkit consists of the following items: 1) a recommendation to request and report vital signs in all pediatric hospital/ED transfers, 2) dissemination of evidence based pediatric sepsis diagnostic criteria to accepting and referring providers and transfer nurses, and 3) accepting physician education in conversational strategies to promote improved diagnostic accuracy during pediatric transfer conversations.

## 2. Study design

This study is a prospective type 2 hybrid implementation-effectiveness trial to evaluate a sepsis diagnostic safety toolkit. We will use explanatory mixed methods to assess quantitative and qualitative metrics, including PRISM contextual factors and RE-AIM outcomes such as accuracy, adoption, adaptations, sustainability, and context.

We will conduct a pre/post-evaluation of the toolkit to measure effectiveness outcomes. We will also evaluate additional implementation outcomes using quantitative and qualitative approaches.

The study will consist of a 24-month pre-intervention period and a 24-month post-intervention period, capturing the same sepsis seasonality during both phases.

## 3. Aims and objectives

The objective of this trial is to implement and evaluate the sepsis diagnostic safety toolkit in an implementation study. We will use explanatory mixed methods to assess quantitative and qualitative metrics, including PRISM contextual factors and RE-AIM outcomes such as accuracy, adoption, adaptations, sustainability, and context.

## 4. Outcomes

### 4.1 Primary effectiveness outcome

The primary outcome will be accuracy, defined as the proportion of included patients in whom all 3 of the following occur while the patient is physically in the referring ED/hospital:

1. Verbal or written use of the term sepsis or septic shock (this may be documented in the chart or stated on the transfer phone call by the referring or accepting physician). Terms that will be considered a diagnosis: sepsis, septic shock, sepsis-y, sepsis-like, septic (referring to the patient). Terms not considered a diagnosis: rule-out sepsis, septic joint, not sepsis.
2. Intravenous antibiotics administered
3. Intravenous fluid bolus administered, or contraindication to administering fluid documented in written or verbal form

Accuracy will be described for important subgroups, including by patient characteristics, ED characteristics, provider characteristics, and time of day, seasonality.

*Sensitivity analysis*: Sepsis is an umbrella term for a set of infectious diagnoses in the setting of organ dysfunction. It is possible to accurately diagnose a specific disease such as pneumonia, and appropriately administer the recommended treatments for sepsis in a timely way, even if the term sepsis has not been used. Thus, we will conduct a sensitivity analysis for our primary outcome of accuracy, in which the diagnosis will be considered accurate if an

infection is diagnosed and guideline-concordant treatments for sepsis are administered, regardless of whether the term "sepsis" is used.

4.2 Implementation outcomes

**Table: RE-AIM Outcomes and PRISM Contextual Factors**

| Component | Outcomes | Analytic Approach |
|---|---|---|
| **Reach** (RE-AIM) | • Number, proportion, representativeness of patients receiving components of sepsis diagnostic toolkit<br>1. Vital signs on transfer call<br>2. The word sepsis used on transfer call | • Descriptive statistics of included population |
| **Effectiveness** (RE-AIM) | • Diagnostic accuracy | • Pre/post proportion difference test<br>• Exploratory: Interrupted time series test of level change |
| **Adoption** (RE-AIM) | • Number of nurses and accepting CHCO physicians using the diagnostic toolkit<br>• Barriers, facilitators | • QUAL: Focus group, periodic reflections<br>• QUANT: Number, proportions, site of transfer center nurses and physicians who use toolkit elements, ICU/PEM |
| **Implementation** (RE-AIM) | • Vital Signs Reported on Phone Call (Any, All)<br>• Duration of transfer calls | • Pre/post proportion difference test<br>• Exploratory: Interrupted time series test of level change<br>• Pre/post comparison of call duration |
| **Maintenance** (RE-AIM) | • Proportion of transfer sepsis patients receiving toolkit in final 3 months vs. first 3 months<br>• Decision of clinical teams to continue toolkit after study period ends | • QUANT: Proportions using toolkit in first vs. final 3 months<br>• QUAL: Decision of clinical, executive leadership to continue use |
| **Recipients** (PRISM) | • Number, site characteristics of accepting CHCO ED physicians<br>• Number of transfer center nurses | • Descriptive tables |
| **Perspectives on the Intervention** (PRISM) | • Acceptability, appropriateness, feasibility of the diagnostic safety toolkit | • QUAL: Interviews of ED providers, focus groups of accepting physicians, transfer nurses' periodic reflections |
| **External Environment** (PRISM) | • External events & policies that interacted with the implementation (e.g. infectious disease, weather, medicolegal) | • PROCESS: Clinical events recorded by study team throughout<br>• QUAL: Focus groups |

4.3 Secondary outcomes

30-Day In-Hospital Mortality or ECMO
Vasoactive agent days during hospitalization
Positive pressure ventilation during hospitalization
Hospital Length of Stay
ICU Length of Stay

4.4 Sample size calculation

We will compare outcomes between pre and post populations, using logistic regression. Pre-specified covariates will be considered for inclusion if they differ substantially (standardized mean difference >0.1) between pre and post models (propensity score weight correction to correct overparameterization). The following covariates have been pre-specified, based on clinical significance and previously reported association of these factors with difficulty in diagnosing sepsis:

-Age
-Fever present in the referring ED/hospital
-Non-Metro Hospital (RUCC 4-9)

The unit of analysis for this study is pediatric patients transferred to CHCO with sepsis (Phoenix score >=2 at the referring hospital, in transit, or within 6 hours of arrival at CHCO, or patients determined by unanimous consensus of 3 independently-reviewing physicians to have had sepsis present at the referring hospital, without appropriate labs measured to meet Phoenix criteria). Based on historic patterns, we expect 100 cases in the pre-intervention period, and 100 cases in the post-intervention period.

Minimal necessary sample size to fit the multivariable models

- Power calculations: We expect to compare 100 cases in the pre-intervention and 100 cases in the post-intervention period. We will use multiple generalized linear models to compare differences in the following 3 outcomes between these two periods:

    - Diagnostic Accuracy: We expect the intervention to increase diagnostic accuracy from 60% pre-intervention to 80% post-intervention. With 200 total patients and alpha = 0.05, we will have a power of 88% to detect this difference.
    - Vital Sign Reporting (Any vital sign): We estimated at baseline that 75% of the study population has vital signs reported, with an estimate of the effect of the intervention to increase reporting that ranges from an absolute increase of 10 to 20%. To detect an increase in vital sign reporting of 20% we will have a power of 99%, while for an increase in vital sign reporting of 10% we will have a power of 43%. Should the treatment effect end up less than an increase of 20%, our minimal detectable increase, fixing power at 80%, is approximately 14.8%.

    - Vital Sign Reporting (All vital signs): We estimated at baseline that 15% of the study population has vital signs reported, with an estimate of the effect of the intervention to increase reporting that ranges from an absolute increase of 20 to 40%. To detect an increase in vital sign reporting of 20% we will have a power of 91%, while for an increase in vital sign reporting of 40% we will have a power of 100%.

- For each comparison above, we have more than sufficient power if the effect size is slightly diminished in the presence of confounders. In fact, with 100 per time period, and a power of 80%, we will have sufficient power to detect a small effect size (0.4) for each comparison. If we divide alpha by 3 to accommodate multiple comparisons, we will still be able to detect an effect size of 0.46.

- Effect size estimates: The existing pediatric sepsis literature to draw on for our estimates of effect sizes varies widely, with differences in population, measures, and hospital characteristics.
  - Diagnostic Accuracy: The literature describing the effects of clinical and quality interventions on improving diagnosis of pediatric sepsis is limited, with widely varied reported pre-, post-accuracy, and interventions reported to increase accuracy by 26-39%. After an intervention, all studies reported accuracy above 50%. The effect of interventions was largest in studies of hospitals with low starting diagnostic accuracy, similar to our expected population.

- Exploratory: We will evaluate an interrupted time series (ITS) model to explore the sensitivity of our results to seasonal factors. Analyses will focus on detection of a **level change** in an outcome, following the intervention, using an ITS design implemented using a logistic regression framework. There will be an equal span of time before and after the intervention.

**4.3 Safety outcomes**

Adverse events will be identified upon occurrence and on data review at the end of the study period.

**5.1 Subgroups**

The primary analysis will be conducted in all patients meeting inclusion criteria.

- We will consider several relevant subgroups including type/size and ruralness of ED, patient age, and medically complex patient. We will assess the intervention in relevant subgroups (representativeness of patients receiving the intervention and providers' adoption, characteristics of the recipients of the intervention). In addition, we will conduct effectiveness evaluations in relevant subgroups, to assess whether the intervention is differentially impacting subgroups. Given the overall sample size, these subgroup analyses are expected to be descriptive, and we will use preliminary statistical analyses with no formal tests of hypotheses, as we do not want to commit type II errors.

**5. Analyses**

- Design: To increase the rigor of the quantitative analysis, and adjust for the potential influence of unbalanced factors, we will assess the quantitative effectiveness measures using multiple generalized linear models (logistic regression for binomial outcomes). These models will include a fixed binary factor for pre- vs. post-intervention and up to 3 additional covariate degrees of freedom determined a priori by the PI and study team

(including age, fever at OSH (y/n), and non-metro hospital (RUCC 4-9)). We will restrict this to 3 due to the expected number of events per binary outcome variable, so as to avoid falling below a parameter to event ratio of 1 per 10 events. This will allow us to control for confounders that vary between the pre- and post-intervention periods.

6. Missing data

Missing data are not expected in this dataset, which only includes existing clinical data on patients through hospital discharge or 30 days, whichever happens first.

Qualitative Analysis Plans

Transfer Calls Conversation Analysis (CA): 50 transfer calls in the post-intervention period will be analyzed using CA, including calls with diagnostic accuracy and with diagnostic missed opportunity. In-depth CA can assess fidelity to healthcare interventions, and enhance education and future implementation based on observed patterns. CA will be used to identify instances of communication practices employed in the delivery and receipt of each of the sepsis diagnosis toolkit components. This analysis will include the presence or absence of each toolkit component, how each component is delivered, and how these variations contribute to diagnosis.

Focus Groups: We will use qualitative coding of fidelity to the intervention to identify high- and low-adopters of the intervention at the accepting physician level in the post-implementation period. Participants will be identified and individually solicited via email, with gift card incentives to encourage participation. One remote, virtual focus group of high-adopters, and one remote, virtual focus group of low-adopters will occur. Participants will not be told that they have been identified as a low- or high-adopter. Virtual groups will have smaller numbers of participants to allow better speaking opportunity in the virtual environment. A semi-structured focus group facilitation guide will be developed including questions about the use of the diagnostic toolkit and perceived outcomes. Acceptability, feasibility, and perceptions of reach and effectiveness will be assessed. Guided by PRISM, multilevel barriers, and facilitators to use of the toolkit will be assessed, as well as expectations about sustainment. We will specifically ask about drivers and barriers to adoption, the interaction between adoption and providers or nurses, organizational (individual ED and health system) characteristics, implementation and sustainability infrastructure, and the external environment.

Periodic Reflection: Periodic, routinely scheduled meetings of the transfer nurse group will be used for periodic reflection on the toolkit. This format allows the study team to hear from all transfer nurses who routinely attend these meetings as a job expectation, thus improving capture of nurses with varied levels of interest in the intervention. There will be immediacy and less risk of recall bias by eliciting information throughout implementation. This will allow the team to understand the experience of the transfer nurses with the intervention. Acceptability, feasibility, and perceptions of reach and effectiveness will be assessed. Guided by PRISM, multilevel barriers, and facilitators to use of the toolkit will be assessed, as well as expectations about sustainment. We will specifically ask about drivers and barriers to adoption, the interaction between adoption and providers or nurses, organizational (individual ED and health system) characteristics, implementation and sustainability infrastructure, and the external environment.

Interviews with Referring Providers: Interviews will be conducted with referring ED providers who have referred children with sepsis after implementation of the diagnostic toolkit. Participants will be identified using the Sepsis Transfer Registry, with purposive sampling to include ED providers from varied types and locations of EDs, those whose calls had diagnostic opportunity or accuracy. Participants will be individually emailed, with participation

incentivized with gift cards. A semi-structured interview guide will be created including questions about diagnosing pediatric sepsis, their experience with the transfer process, the local hospital environment and culture, and needs and wishes for the transfer and diagnostic support experience. Guided by PRISM, multilevel barriers and facilitators to pediatric sepsis diagnosis will be assessed to identify any gaps in support provided by consultative calls with the CHCO transfer call center using the toolkit.

Data Collection & Analysis: Trained qualitative interviewers will conduct the focus groups and interviews, which will be audio recorded, transcribed then checked for accuracy. Data from the focus groups and interviews will be coded using Atlas.ti. Two members of the study team will independently review transcripts, notes, and summary sheets and identify codes. Interviews will be conducted until thematic saturation is achieved, which we will assess when no new content is identified for 3 interviews in a row, and interviews are repetitious. Codes and coded transcripts will be collectively reviewed by the research team who will reconcile discrepancies. A thematic codebook will be established and applied to the transcript. Codes will be clustered into related groups to guide development of themes. Deductive analysis guided by the PRISM framework will be used to address study aims.

Data Integration**:** The quantitative data will be used to help select participants for the qualitative component. Data will be combined for discussion and contextualization of findings.

## 7. References

James Lopez Bernal, Steven Cummins, Antonio Gasparrini, Interrupted time series regression for the evaluation of public health interventions: a tutorial, *International Journal of Epidemiology*, Volume 46, Issue 1, February 2017, Pages 348–355, https://doi.org/10.1093/ije/dyw098

Hategeka C, Ruton H, Karamouzian M, Lynd LD, Law MR. Use of interrupted time series methods in the evaluation of health system quality improvement interventions: a methodological systematic review. *BMJ Glob Health*. 2020;5(10):e003567. doi:10.1136/bmjgh-2020-003567

Saeed, S., Moodie, E.E.M., Strumpf, E.C. *et al.* Segmented generalized mixed effect models to evaluate health outcomes. *Int J Public Health* **63**, 547–551 (2018). https://doi.org/10.1007/s00038-018-1091-9