# Predicting Inpatient Pediatric Outcomes (PIPO)

**Derek J. Williams, MD, MPH**

**Department of Pediatrics**
**Vanderbilt University School of Medicine**

**Hank Domenico**
**Yuwei Zhu**
**Dan Byrne**

**Department of Biostatistics**
**Vanderbilt University School of Medicine**

**Table of Contents:**

## 1.0 Background

In medicine, clinical prediction rules (CPRs) are used to assist providers in decision-making by standardizing clinical assessments, improving the accuracy of risk estimates, and tailoring management decisions. Using principles of modern statistical modeling, CPRs aggregate data elements from a variety of sources (e.g. demographic, clinical, laboratory, and/or radiographic factors) to predict outcomes. Several high quality CPRs have been developed and broadly validated for use in pediatric populations.

## 2.0 Rationale and Specific Aims

The purpose of this study, Predicting Inpatient Pediatric Outcomes (PIPO), is to use retrospective data on children hospitalized at Vanderbilt to develop clinically meaningful CPRs for a variety of different outcomes. Our intent is to use retrospective data available through Vanderbilt databases (e.g. electronic data warehouse [EDW]) such that well-performing CPRs can eventually be incorporated into automated electronic decision support tools. After initial development and thorough validation, we will prospectively deploy well-performing CPRs in randomized trials (following additional IRB approval).

The primary objective of PIPO is to develop clinically-relevant CPRs for use among hospitalized children using data from Vanderbilt's rich array of electronic databases, primarily the EDW. Examples of planned projects include CPRs for predicting acute clinical deterioration (e.g. unplanned transfer to intensive care), hospital readmissions, and healthcare and non-healthcare associated infections (e.g. community-associated MRSA disease, central-line associated blood stream infections, and *Clostridium difficile* enteritis).

## 3.0 Previous Studies

This group has previously engaged in similar projects for adults including predicting risk of readmission, pressure ulcers, and other adverse outcomes. Variables routinely collected in the electronic health record were used to build predictive models of each patient's risk of these events. After validation, these predictive models were built into the electronic health record and are being used to automatically calculate each patient's probability of developing these outcomes. Interventions are currently being designed to focus resourses on patients at the highest risk.

## 4.0 Inclusion/Exclusion Criteria

- Children < 18 years of age

- Hospitalized at The Monroe Carell Jr. Children's Hospital at Vanderbilt

## 5.0 **Enrollment/Randomization**

The study will only utilize retrospective data collected routinely for non-research purposes between January 2009-December 2015. Only data reliably reported to Vanderbilt's local database resources will be used. Due to the study's retrospective nature as well as the anticipated large number of hospitalizations (~13-14,000 annually, approximately 100,000 total) and minimal risk to human subjects, a waiver of consent will be obtained.

## 6.0 **Study Procedures**

Data collected for PIPO will be obtained exclusively through Vanderbilt's electronic databases, such as the EDW, Research derivative, Quality Safety and Risk Prevention databases, etc . The following types of data will be abstracted from medical records of children hospitalized at Monroe Carell Jr. Children's Hospital at Vanderbilt:

- Admission month and year
- Sociodemographics (e.g. age, sex, race/ethnicity, insurance status)
- Zip code (for determination of distance from hospital and area income status)
- Hospital location (e.g. acute vs. critical care, service)
- ICD9/CPT coded diagnoses/procedures
- Comorbidities (e.g. cardiopulmonary, neurologic, genetic disorders, etc.)
- Prior and/or subsequent hospitalizations at Vanderbilt
- Vital signs (e.g. temperature, heart rate, respiratory rate, oxygen use, etc.)
- Treatments and procedures (e.g. medications, respiratory care, central line insertion, etc.)
- Laboratory and microbiology data (e.g. electrolytes, blood counts, viral antigen testing, etc)
- Microbiology data (e.g. organisms and drug susceptibilities)
- Radiology studies
- Adverse events to be used as outcomes in our predictive models

Note, each CPR will be unique and not all data elements above will be utilized for every CPR developed.

Protected health information (with the exception of elements of dates and zip code information) will only be used to identify medical records for review and will not be included in the development of the proposed CPRs. Furthermore, personal identifiers used to access records will be stored securely, on a password-protected computer

accessible only to key study personnel and only for the minimum amount of time necessary to access the medical records and complete the study (1 year after publication of peer-reviewed manuscripts), after which time the personal identifiers for the corresponding record will be destroyed.

The goal of the study is to model the risk of adverse events based on historical data, then deploy the models in real time, using the electronic health record, to stratify patients and intervene on those with the highest risk. Aim one will be to obtain historical data and use it to build predictive models of the adverse outcomes listed above. The predictive models will be logistic regression, survival modeling, or continuous regression depending on the form of the outcome and relevant predictors. The most recent year of data will be set aside for model validation. Following separate IRB approval, well performing models will be built into the electronic medical record and used to identify patients at high risk of adverse outcomes.

## 7.0 Risks

As a retrospective study using electronic data sources, there is no risk to human subjects. To maintain confidentiality, data will be accessible only to key study personnel on their computers.

## 8.0 Reporting of Adverse Events or Unanticipated Problems involving Risk to Participants or Others

All key study personnel are trained in the conduct of human subjects research and have particular experience with the types of studies outlined in this application. Data will be stored electronically via secure, password-protected servers available only to key study personnel. The study team will meet monthly to review study progress and will hold ad hoc meetings if any issues or unforeseen events arise. We do not anticipate these studies will result in any AEs. Any member of the study team who suspects possible loss of confidentiality or other unforeseen risk to human subjects will notify the study PI immediately. The PI and study team will immediately halt all study related activities and notify the IRB of any such events, while also attempting to identify root cause of the breach of study protocol and planned intervention to minimize future risk.

## 9.0 Study Withdrawal/Discontinuation

We do not anticipate study withdrawal or discontinuation as the study is retrospective using data previously collected at Vanderbilt for non-research purposes.

## 10.0 Statistical Considerations

Risk models will be developed using standard, logistic, or Cox regression depending on the form of the outcome variables and predictors. For each model, potential predictor variables will be selected on the basis of existing literature, expert input, and frequency of outcome. Missing data will be handled using complete case analysis, multiple imputation, or separate modeling in different patient groups depending on the pattern of missingness. Internal bootstrap calibration will be performed on each model as well as external validation on one year of data set aside from the modeling process.

## 11.0 **Privacy/Confidentiality Issues**

We plan to use data available exclusively through Vanderbilt's electronic resources (e.g. EDW). All data will be collected retrospectively. No prospective data collection or human subject interactions are planned.

As a retrospective study using electronic data sources, there is no risk to human subjects. Evaluation reports will use group summaries and no PHI will be published or made available to non-study personnel. To conduct this study we need to be able to merge data sets and verify results. Therefore the files cannot be deidentified.

All key study personnel are trained in the conduct of human subjects research.

## 12.0 **Follow-up and Record Retention**

Data collected for this study will be obtained exclusively through Vanderbilt's electronic databases, such as the EDW, Research derivative, Quality Safety and Risk Prevention databases, etc . Protected health information (with the exception of elements of dates and zip code information) will only be used to identify medical records for review and will not be included in the development of the proposed CPRs. Furthermore, personal identifiers used to access records will be stored securely, on a password-protected computer accessible only to key study personnel and only for the minimum amount of time necessary to access the medical records and complete the study (1 year after publication of peer-reviewed manuscripts), after which time the personal identifiers for the corresponding record will be destroyed.