



Jan 23,  
2020

## Assessment of safety and effectiveness for pediatric patients of cryoanalgesia compared to epidural for the Nuss procedure in real-world data from Electronic Health Records: a protocol and statistical analysis plan

Benjamin S. Glicksberg<sup>1,2,3</sup>, Roberto Mora<sup>4</sup>, Stefano Rensi<sup>4,5</sup>, Vivek A. Rudrapatna<sup>1</sup>, Andrew M. Bishara<sup>1,6</sup>, Elizabeth Gress<sup>4,7</sup>, Rohit Vashisht<sup>1</sup>, Michael Zobel<sup>4</sup>, Benjamin Padilla<sup>4,8</sup>, Benjamin Padilla<sup>4,8</sup>, Atul J. Butte<sup>1</sup>

<sup>1</sup>Bakar Computational Health Sciences Institute, University of California, San Francisco, USA, <sup>2</sup>Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, USA, <sup>3</sup>Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA, <sup>4</sup>Department of Surgery, University of California, San Francisco, USA, <sup>5</sup>Department of Bioengineering, Stanford University, USA, <sup>6</sup>Department of Anesthesia and Perioperative Care, University of California San Francisco, San Francisco, California, USA, <sup>7</sup>Department of Bioengineering & Therapeutic Sciences, University of California, San Francisco, USA, <sup>8</sup>Department of Surgery, Phoenix Children's Hospital, USA

**1** Works for me [dx.doi.org/10.17504/protocols.io.bbpaime](https://doi.org/10.17504/protocols.io.bbpaime)

Benjamin Glicksberg  
University of California, San Francisco

### ABSTRACT

Randomized control trials (RCTs) are the gold standard for clinical trials to evaluate effectiveness and safety of interventions. Conducting RCTs, however, is not always feasible, particularly for rare conditions. Real-world data (RWD), including clinical information from Electronic Health Records (EHR), are utilized for similar analyses in retrospective data. The United States Food and Drug Administration are increasingly receptive of accepting such analyses as evidence as long as the studies are designed and conducted robustly. EHR work suffers from a reproducibility crisis due to such factors such as hospital-specific biases, among others. Akin to preregistration of RCTs, devising, detailing, and publishing protocols of RWD before conducting analyses can enhance reproducibility and fidelity of results. In the current work, we release a detailed protocol to assess effectiveness and safety of a cryoablation intervention compared to epidural for pain control in adolescent patients undergoing the Nuss procedure in the retrospective EHRs of a major hospital system.

### GUIDELINES

- This work attempts to extend the scope of a Randomized Control Trial (RCT) comparing the effect of cryoablation vs. epidural in Nuss procedure in terms of safety and effectiveness (Graves et al., Intraoperative intercostal nerve cryoablation During the Nuss procedure reduces length of stay and opioid requirement: A randomized clinical trial. *J Pediatr Surg*, 54, 2019). The authors identified significant benefit of the cryoablation procedure in such variables such as length-of-stay to discharge and pain.
- This work has been funded by the U.S. Food & Drug Administration as part of a Real-World Evidence Demonstration Project under Pediatric Device Consortia grant no. 1P50FD006424-01 ("UCSF-Stanford Pediatric Device Consortium," PI: Michael Harrison). Neither the UCSF-Stanford PDC nor any of the protocol authors has a financial relationship or other conflict of interest with AtriCure, Inc., the manufacturer of the cryoablation device studied in this analysis.
- All patient data were obtained from the University of California, San Francisco (UCSF) healthcare electronic health records (EHR) system. Our data spans the years 2013 to 2019. UCSF uses APeX implementation of EPIC. In this protocol, we outline the steps used to collect information within this system.
- For many retrospective studies, selection bias based on clinical assessment is difficult to identify, quantify, and address. In other words, was the decision of what procedure to perform related to the potential outcome? In this study, however, there was a clear transition of clinical practice to exclusive use of the cryoablation procedure in all patients following completion of the aforementioned RCT. Thus, at a given time point, all patients received the identical intervention, eliminating the potential selection bias as a confounding variable.
- At the time of the publication of this protocol, the process of data extraction and abstraction had already begun but was not yet completed.

### MATERIALS TEXT

- An encrypted computer to handle all data processing.
- A HIPAA-compliant, secured server to house and process identified patient data
- Access to Electronic Health Record system data: this project is built around *Epic*<sup>TM</sup> software.
- Data processing and statistical analysis software (e.g., *R*)

#### SAFETY WARNINGS

- All researchers adhered to strict HIPAA compliance and have up-to-date training or certification in human subjects protection ([www.CITIprogram.org](http://www.CITIprogram.org)).
- All computers that were used to store and/or view these data were properly encrypted according to UCSF's standards.

#### BEFORE STARTING

- The project, which involves access to identified EHR data, was approved by the UCSF IRB.
- For this project, a pre-selected list of patient identifiers was provided for individuals having undergone Nuss procedure at UCSF.

### Cohort Identification of Nuss Procedure

- 1 The pediatric surgical department provided a list of medical record numbers of all patients that underwent a Nuss procedure. We will only analyze data for patients that were under 22 years of age (i.e., pediatric population) at the time of procedure. For all patients, we checked all "Admission" encounters to verify that a Nuss procedure was actually performed. From here, we checked under "Brief Hospital Course by Problem" for a mention of Nuss procedure. This was often described explicitly, but was sometimes detailed as "pediatrics laparoscopic repair of pectus excavatum" or "thoracoscopic repair of pectus excavatum". In all cases of how the procedure was labeled, we further verified that it was the Nuss procedure being performed by checking the "Care Timeline" field in the admission report by selecting the procedure of interest and reading the details for confirmation of both "Nuss bar placement" and "repair of pectus excavatum". Alternatively, confirmation of these elements could also be found under the "Procedure Performed and Complications" section. This identified encounter was used as the starting point for other variable collection strategies.

### Variable Selection

- 2 The following subsections detail the procedures of collecting and storing relevant variables that are used for analyses from our EHR system.

#### 2.1 Cryoablation vs. epidural intervention verification

Due to potential inconsistencies in documenting cryoablation vs. epidural procedures, particularly due to the fact some patients were part of a clinical trial, we manually verified these labels. First, we checked the "Admission" encounter under the "Chart Review" tab for the Nuss procedure encounter (identified above), specifically in the "Brief Hospital Course by Problem" field. Here we searched for iterations of "thoracic epidural" or "epidural" placement vs. specific mentions of "cryoanalgesia" or "cryoablation". In order to further verify the indication, we navigated to "Detailed Report" section where we checked all relevant notes during the course of the hospital stay under the "All notes" section. Here we read for explicit mention of cryoablation of the intercostal nerves intraoperatively. If no such indication was available, this suggested that an epidural was placed. In order to corroborate epidural placement, we identified the specific location of where the epidural was placed.

#### 2.2 Gender {M,F,NA}

Gender information was gathered from the specific EHR record of the procedure (see above for details). Specifically, under the "Admission" encounter, there is a "Patient Information" field which lists the patient's gender.

### 2.3 Ethnicity {standardized text, NA}

Self-reported ethnicity information was obtained from the demographic face sheet or de-identified data. The patient information table we obtained, captured the ethnicity field for all patients. Each health system may record the specific values for ethnicity differently based on specific ontological implementation. The two most prevalent known ethnicity options in our data were "Not Hispanic or Latino" and "Hispanic/Latino". As with current best practices, ethnicity and race denominations should be treated separately. We combined similar values, especially for unknown-related options together, such as "Unknown" and "Declined". In a similar vein, it was also necessary to combine infrequent values into an "Other" category for privacy and/or statistical considerations. If there were ever multiple entries for a single patient, a) if there was one unknown-related value and a declared value, we selected with the declared one and b) if there were multiple that were declared and conflicting, we marked as "Other".

### 2.4 Race {standardized text, NA}

Like self-reported ethnicity information, self-reported race information was obtained from the demographic face sheet or the de-identified data. In the patient race information table, which was separate from the primary patient table, we obtained race information for all patients. Each health system may record the specific values for race differently based on specific ontological implementation. In our version, there were fields such as "White or Caucasian" and "Black or African American". As with current best practices, ethnicity and race denominations should be treated separately. It may be necessary to combine similar values, especially for unknown-related options, together, such as "Unknown" and "Declined". In a similar vein, it also may be necessary to combine infrequent values into an "Other" category for privacy and/or statistical considerations. If there were ever multiple entries for a single patient, a) if there was one unknown-related value and a declared value, we selected with the declared one and b) if there were multiple that were declared and conflicting, we marked as "Other".

### 2.5 Date of birth {YYYY-MM-DD, NA}

Date of birth for patients was gathered from APeX in the specific record of the procedure. Specifically, under the "Admission" encounter, there is a "Patient Information" field which lists the patient's date of birth.

### 2.6 Date of surgery {YYYY-MM-DD, NA}

Date of surgery (DOS), or alternatively date of procedure, for patients was gathered from Apex in the specific record of the procedure. At the bottom of the "Admission" encounter, the "Detailed Report" indication was selected followed by "All Notes". These notes were checked to identify the DOS. The fields "Chief Complaint and Brief HPI" and "Brief Hospital Course by Problem" were checked to verify the procedure performed was the correct one (i.e., Nuss) as well as the reported age at surgery, which was compared to our calculation from the patient's DOS and date of birth.

### 2.7 Operating surgeon {Free text: name, NA}

While the information pertaining to operating surgeon will be obtained as an identifiable name, we recommend recording it as a unique, synthetic ID for privacy concerns. This information was gathered per patient's surgery from Apex in the specific record of the procedure. In the top left corner of the "Admission" report right above the "Last attending" indication, information pertaining to the surgeon or surgeons performing the procedure. In cases where there were multiple operating surgeons, both should be recorded as the statistical approaches we will employ can account for it.

## 2.8 Haller index {Numeric (float), NA}

To obtain the most up-to-date Haller index for each patient, we selected the chart review tab, then the encounters tab, and then the "Admission" encounter for the designated date of surgery (identified via Nuss procedure variable description above). In cases where the Haller index was available from this encounter, it was recorded from within the "Chief Complaint and Brief HPI" field. If not available under this field, the score was then searched for by selecting "Detailed Report" within the "All Notes" field.

If the Haller index score was not available via these steps, we checked the specific surgery and anesthesia encounters for the date of surgery. If not available in these encounters, then all prior encounters before the "Admission" encounter were checked for the most recent (relative to admission date) Haller index. In these scenarios, we were generally able to obtain the score in one or two prior encounters.

In these scenarios, this information was most commonly listed under "Office Visit" encounters in which the provider would mention the Haller index from a CT under the field "History of Present Illness" although for some other office visits the Haller index could be found under the "Progress Notes" field or under the "Studies" field found on an office visit.

This information was next second most commonly listed under the "Appointment" or "CT Chest without contrast" encounter with the label "RAD CT." In this type of encounter, looking in the "Orders Performed" field and selecting the "CT Chest Without Contrast" often clearly lists the Haller index.

In the rare case where two Haller index scores were present in the same note, we selected the score given by the operating surgeon when available or most recent and/or the most recent score otherwise. In other rare cases, Haller index may be obtained from "Telephone" encounter documentation prior to the "Admission" encounter. In extremely rare cases where the Haller index could not be identified by any of the above steps, the PI of the study helped to characterize scores using the CT scans available in Apex to retrospectively calculate a Haller index at the date of surgery for patients. In these scenarios, where Haller index was not obtained on date of procedure, we recorded the amount of time between the procedure date and last Haller index recording that was used.

## 2.9 Revision surgery? {Yes, No, NA}

To determine whether the procedure was a revision or not, we navigated to the specific surgical encounter (identified through steps listed above), then "Detailed Report", then "All Notes". Here is where information on any prior related procedures is detailed.

## 2.10 Number of prior related procedures {Numeric (float), NA}

For information pertaining to prior related procedures (e.g., Ravitch) obtained in Step 10, we tabulated this amount and added as a numeric value. For instance, this variable would be a "2" if a patient had two prior related procedures before the one of interest.

## 2.11 Body-Mass Index (BMI) {Numeric (integer), NA}

BMI was gathered from "Admission" encounters under the Apex system chart review. However, for a few patients' "Admission" encounters, height and/or weight was not collected and thus no BMI was calculated. As such, we navigated the most proximal previous or subsequent encounters from the date of admission up to two weeks in either direction to identify BMI or calculate through independent weight/height variables.

### Outcome/Endpoint Variable Collection

- 3 The following subsections detail the procedures of collecting and storing endpoint-related variables from our EHR system.

### 3.1 Length of stay (post-op days) {Numeric (integer), NA}

Length of stay (LOS) was determined by entering the chart review tab going to the "Admission" encounter for the relevant Nuss procedure, which details both admission and discharge dates. The total LOS was calculated based off these dates. We will focus on LOS in terms of Post-Op Day (POD), specifically the number of days spent in the hospital after surgery was performed. This decision was made in order to protect against possible scenarios where patients may have spent some time in the hospital before the surgery.

### 3.2 Pneumothorax {Yes, No, NA} and {YYYY-MM-DD, NA} and {free text, NA}

Information regarding development of pneumothorax was obtained from selecting the "Detailed Report" under the "Admission" encounter for the relevant procedure (steps to identify this listed above) and then selecting the "All Notes".

Development of pneumothorax was not always clear based off these notes, and as such the "Imaging" tab was selected in order to more accurately record these variables. From the surgery to discharge period, we read "Impression" and "Findings" fields for all "X-ray Encounters" within the admission to discharge period for any and all mentions of pneumothorax, which were recorded for endpoint classification. We recorded relevant information as free text such as whether the pneumothorax resolved on its own. The specific characterization of clinically-relevant pneumothorax includes information from this and the Chest Tube Insertion sections and is described in the Endpoint section below.

### 3.3 Chest tube insertion {Yes, No, NA} and {YYYY-MM-DD, NA} and {free text, NA}

Information regarding chest tube insertion was obtained from selecting the "Detailed Report" under the "Admission" encounter for the relevant procedure (steps to identify this listed above) and then selecting the "All Notes". Insertion of chest tube was not always clear based off these notes, and as such the "Imaging" tab was selected in order to more accurately record these variables. Here, we read "Impression" and "Findings" fields for all "X-ray Encounters" within the perioperative period until 30 days post-discharge for any and all mentions of chest tube insertion, which were recorded for endpoint classification, specifically detailing if the procedure was done during the initial surgical encounter recovery period or in a readmission. The specific characterization how this variable was used to define the outcome of interest, specifically clinically-relevant pneumothorax, is described in the Endpoint section below.

### 3.4 30-day readmissions {Yes, No, NA} and {YYYY-MM-DD, NA} and {free text, NA}

We identified patient readmissions to the hospital from discharge to 30-days after by assessing all subsequent encounters during this period. Specifically, we looked for encounters under "Chart Review". All readmissions encounters were clearly indicated by bold red letters. The specific labeling of these encounters differed, but the majority were "ED" or "Admission". We also recorded free text from the notes that described the reason for the readmission, specifically relevant procedure and diagnosis. The locations of where these pieces of information were found varied by encounter type, but most often, they could be found under "Brief Hospital Course by Problem" and further diagnosis-related information could be found in the "Diagnosis" field for the "Admission".

## Data Quality Control Processes

4 The following section details various control processes we perform to address data quality.

## 4.1 Missing data points

Many times, in EHR-based research that does not have a prospective study design, important data are missing for multiple variables. For some variables, this uncertainty is embedded in the collection process, such as “Unknown” for demographic variables. There are various strategies to deal with this missingness such as creating an “Unknown” option (i.e., for categorical variables), imputing (i.e., primarily for numerical variables), or removing. This decision of how to deal with missingness revolves around considerations of the particular question as well as scope and quality of the dataset (among many others). Due to the small sample size, we opted not to perform imputation for missing variables.

For this project, we will include “Unknown” option for demographic variables, but not include any patients in specific analyses that are missing key endpoints (i.e., LOS). More specifically, a patient would only be excluded for the specific analysis in which there is no information but included for the ones in which there are data available. We will also not include any patients that have incomplete information that is required for analysis (e.g., missing co-variate for BMI). Regarding missing covariate information, we have developed strategies to acquire information from other time points that are akin to some imputation strategies (i.e., carry last observation forward; also, see Endpoints section).

## Study Design

5 The following section details specifics of the current study design.

### 5.1 Effectiveness Hypotheses

Null hypotheses:

There is no significant difference in effectiveness of cryoablation compared to epidural for Nuss procedure.

Alternative hypotheses:

There is a significant difference in effectiveness outcome for cryoablation during the Nuss procedure compared to epidural.

### 5.2 Safety Hypotheses

For the safety endpoints, we will report results using descriptive statistics. While we will also test the hypotheses listed below, our conclusions will be limited and cannot go beyond the ability to reject or fail to reject the null hypothesis. Specifically, if we reject the null hypothesis, we cannot confidently say that the Cryo procedure is as safe as epidural overall (i.e., equivalence), just that we failed to see a difference in our particular sample.

Null hypothesis:

There is no significant difference in safety of cryoablation compared to epidural for Nuss procedure.

Alternative hypothesis:

There is a significant difference in safety outcomes of cryoablation compared to epidural for Nuss procedure.

### 5.3 Sample size considerations

This is a study of retrospective data and requires no prospective collection. We performed a power calculation based on the results of the original RCT study of interest (Graves et al., Intraoperative intercostal nerve cryoablation during the Nuss procedure reduces length of stay and opioid requirement: A randomized clinical trial. *J Pediatr Surg*, 54, 2019). For this power calculation, we utilized the following parameters: Mann-Whitney-U statistical test, two tailed experiment, normal parent distribution,  $\alpha = 0.05$ ,  $\beta = 0.1$ , power = 0.9, and allocation ratio = 1. The LOS results from the prior study were implemented as such: mean of group 1 = 4.9, mean of group 2 = 2.8, and standard deviation across both groups = 1.4, which produced a Cohen's d (effect size) of 1.5. Entering these values to calculate sample size produced the following results: total sample size = 22, actual power = 0.9, degrees of freedom = 19, critical t = 2.1, and noncentrality parameter ( $\lambda$ ) = 3.4. Therefore, we would need at least 22 samples per group to achieve statistical power. We performed all calculations using the G\*Power (version 3.1) software.

## 5.4 Relevant covariate selection procedure

On the basis of the published literature and clinical experience, we selected the following covariates *a priori* as being effect modifiers of (i.e. having statistical dependencies on) the outcome variables: age, gender, race, BMI, ethnicity, Haller index, days since Haller index recorded, number of prior related procedures, and surgeon. In order to assess the possibility of overfitting, we will perform a sensitivity analysis for each model. Specifically, we will assess goodness of fit via adjusted  $R^2$ , AIC, and BIC and compare against the model without any covariates selected. As a follow-up step, we will also do backward stepwise variable selection to determine the optimal variables to choose.

### Endpoints

- 6 Each safety and effectiveness endpoint(s) will be assessed between cryoablation and epidural groups within pediatric population only (<22 years of age).

#### 6.1 Primary Effectiveness Endpoint(s)

##### *Length of stay*

Is there a significant difference in length of hospital stay (i.e., perioperative period) between the two groups? This is defined as the amount of post-operative days from surgery to discharge from hospital.

This will be assessed via a Poisson regression model (if the estimated variance is proportional to the expected values) or a log-transformed linear regression (if the residual variance is proportional to the squared expected values) with numeric LOS as outcome and intervention as the main variable of interest. This analysis will be unadjusted due to a single hypothesis being tested. All relevant covariates will be included according to the "Relevant covariate selection procedure" described above.

#### 6.2 Primary Safety Endpoint(s)

##### *30-day Readmissions*

Is there a significant difference between groups in the proportion of 30-day hospital readmissions after discharge for a related or relevant issue (i.e., admission to the hospital within 30 days of discharge)?

All patients will be annotated as to whether they had readmission to the hospital within the 30-day period after initial discharge using a binary outcome variable (0 or 1). We will perform a logistic regression using this generated binary variable as the outcome with the intervention being the primary variable of interest. This analysis will be unadjusted due to a single hypothesis being tested. All relevant covariates will be included according to the "Relevant covariate selection procedure" described above.

##### *Incidence of Clinically-Relevant Pneumothorax*

Is there a significant difference between groups in the development of clinically relevant pneumothorax during the perioperative period up to a 30-day window?

Small pneumothoraces can occur during the cryoablation procedure, but many resolve on their own. As such, we will assess safety of cryoablation procedure in whether some action, specifically inserting a chest tube, was required to address the pneumothorax by the treating physician after original surgery. Therefore, a clinically relevant pneumothorax was defined as a patient having a record of a pneumothorax, which required insertion of a chest tube in a procedure subsequent to the original one within 30 days post-op.

For these analyses, we will perform a logistic regression with binary 0 or 1 response corresponding to whether pneumothorax with chest tube insertion occurred with the intervention being the primary variable of interest. This analysis will be unadjusted due to a single hypothesis being tested. All relevant covariates will be included according to the "Relevant covariate selection procedure" described above.



## 7 The following section will detail specifics regarding the statistical analysis plan.

### 7.1 Baseline cohort characteristic differences

As this study is not a randomized control design, it is especially important to verify there are no inherent differences in the two cohorts that might bias the interpretation of the outcome analyses. We will tabulate all demographic and relevant characteristics into a cohort table, specifically: age, gender, ethnicity, race, Haller index, number of prior related surgeries, and body mass index. We will compare the means of numeric data using a two-tailed t-test if the data are normally distributed and Wilcoxon-Mann-Whitney if not, and categorical variables using a chi-square test. This analysis will be unadjusted due to a single hypothesis being tested. If there are significant differences in basic characteristics between groups, we will ensure that the variable is included as a covariate in the model.

### 7.2 Assessment of missingness

As we will not be including patients with key missing variables for the analyses that incorporate such information, there is a major concern of underlying bias of whether these are missing by random chance or not. As such, we will perform Little's test to assess whether there is some other factor that relates to why these data are missing which might bias results. If there are significant differences between the groups, we would report it as a limitation and consideration for interpretation.

### 7.3 Re-analyzing endpoints for RCT cohort using retrospective data

As a preliminary approach, we will re-analyze the original RCT cohort (Graves et al., Intraoperative intercostal nerve cryoablation During the Nuss procedure reduces length of stay and opioid requirement: A randomized clinical trial. *J Pediatr Surg*, 54, 2019) for all overlapping available endpoints using retrospective EHR data, specifically the LOS. We will perform this analysis using both the method of the original study, specifically the Mann-Whitney U-test for nonparametric continuous variables, as well as the procedure detailed in the primary endpoint section. The purpose of this step is to determine the quality of data collected in two varying procedures (i.e., prospective vs. retrospective).



This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited