

Heinz 95-845: Project Proposal

Xinmi Li

Heinz College

Carnegie Mellon University

Pittsburgh, PA, United States

XINMIL/XINMIL@ANDREW.CMU.EDU

1. What is your proposed analysis? What are the likely outcomes?

This study is proposed to provide a retrospective EHR-data-driven examination and analysis of a prospective randomized controlled trial (referred as reference study below), on the effects of stroke volume-guided intravenous fluid therapy and low-dose dopexamine on tissue microvascular flow and oxygenation and systemic markers of inflammation in patients admitted to ICU following gastrointestinal surgery, through machine learning models.

2. Why is your proposed analysis important?

Complications are common after major non-cardiac surgery and represent an important cause of avoidable morbidity and mortality. With around 234 million major surgical procedures performed worldwide each year, around 15% of patients are at high-risk of complications, with mortality rates of up to 12%, and they account for over 80% of early post-operative deaths and a significantly reduced long-term survival rate.

Some relatively small studies have suggested the association between the use of flow related end-points for intravenous fluid and/or low-dose inotropic therapy and improved tissue perfusion and oxygenation, which may reduce the incidence of complications and organ dysfunction. Therefore, in order to improve the clinical outcomes for the post-operative patients, it is important to find the mechanism underlying this benefit and provide a rational basis to guide the uses of such therapies in ICU.

3. How will your analysis contribute to existing work? Provide references.

The reference study has testified the benefits on improving global oxygen delivery, microvascular flow and tissue oxygenation by using stroke volume-guided intravenous fluid therapy and low-dose dopexamine through the randomized controlled trial in UK. Compared with the reference study, this analysis will examine and improve its conclusions through machine learning algorithms which allow a smarter feature selection and model development. The data used in this analysis is from MIMIC II Clinical Database, which has a broader longitudinal scope and larger datasets, collected from a variety of ICUs between 2001 and 2008. Through machine learning, probable bias caused by human factors in the reference study can be reduced, and other subgroup characteristics that contribute significantly to outcomes may be discovered.

4. Describe the data. Please also define Y outcome(s), U treatment, V covariates, W population as applicable.

Y: The primary outcome of this study is the instances of predefined in-hospital complications. The second outcomes are acute kidney injury within seven days, mortality and duration of hospital stay.

U: Three haemodynamic protocols are allocated in this study: Intravenous lactated Ringers solution was administered at 1 ml/kg/hr to maintain a sustained rise in 1) central venous pressure (CVP

group) and 2) stroke volume (SV group), and 3) stroke volume guided fluid therapy was combined with dopexamine (0.5 mcg/kg/min) (SV & DPX group).

V: The covariates are age, sex, and surgical procedure (upper gastrointestinal surgery, lower gastrointestinal surgery and pancreatic surgery involving the gut).

W: The eligible population are patients scheduled for admission to critical care following major elective gastrointestinal surgery, with the exclusions of pregnancy, receiving palliative treatment only and acute arrhythmias or myocardial ischaemia prior to enrollment, receiving lithium therapy, and with a body mass less than 40 kg.

5. What evaluation measures are appropriate for the analysis? Which measures will you use?

The ROC curve, PR curve, sensitivity and specificity. In this analysis, sensitivity will be used, because the purpose of the reference study is to detect the high-risk patients in early post-operation and prevent complications. Therefore, the rate of true positive predictions outweighs other evaluation measures.

6. What study design, pre-processing, and machine learning methods do you intend to use? Justify that the analysis is of appropriate size for a course project.

Given the purpose of this analysis, features selected are according to the reference study. For the missing values, they are imputed based on maximum likelihood estimation. If there are missing features that could be used to build the model, correlated/relative features will be selected with reference on medical knowledge. In this analysis, supervised learning algorithms such as logistic regression, naive Bayes, tree augmented Bayes, and decision tree will be used. Their outcomes will be evaluated and the best model will be used to be compared with the reference study outcome.

7. What are possible limitations of the study?

First, MIMIC II data is collected from one single tertiary teaching hospital. Although the dataset used in this analysis is larger, it has no advantage of patient variety over the reference study. Second, some data elements from MIMIC II Database are represented as notations rather than numbers, such as ranges. In some cases they need to be changed into actual numbers to develop the model. The transformation of the data can introduce errors to the analysis. Third, features from MIMIC II Database and the reference study are diagnostic-dependent and coded under ICD-9, which is now out-of-date and replaced by ICD-10. The conclusion drawn from the analysis may not accurate enough to be applied to today's clinical decision-making.

References

Jhanji S, Vivian-Smith A, Lucena-Amaro S, Watson D, Hinds CJ, Pearse RM. Haemodynamic optimisation improves tissue microvascular flow and oxygenation after major surgery: a randomised controlled trial. *Critical Care*. 2010;14(4):R151. doi:10.1186/cc9220.