

# The Value of Missing Information in Severity of Illness Score Development

**Osman Özaltın, PhD**

Assistant Professor

Edward P. Fitts Department of  
Industrial and Systems Engineering  
North Carolina State University

**Abstract:** The aim of this study is to investigate the hypothesis that using information about which variables are missing along with appropriate imputation improves the performance of severity of illness scoring systems used to predict critical patient outcomes. We quantify the impact of missing and imputed variables on the performance of prediction models used in the development of a sepsis-related severity of illness scoring system. Electronic health records (EHR) data were compiled from Christiana Care Health System on 119,968 adult patients hospitalized between July 2013 and December 2015. Two outcomes of interest were considered for prediction: (1) first transfer to intensive care unit (ICU) and (2) in-hospital mortality. Five different prediction models were used. Indicators were utilized in these prediction models to identify when variables were missing and imputed. We observed a significant increase in prediction performance when moving from models that did not indicate missing information to those that did. Moreover, this increase was higher in models that use summary variables as predictors compared to those that use all variables. We conclude that indicators for missing variables should be incorporated along with appropriate imputation when developing prediction models using EHR data.

**Bio:** Osman Özaltın is an Assistant Professor in the Edward P. Fitts Department of Industrial and Systems Engineering, and a member of the Personalized Medicine Faculty Cluster at the North Carolina State University. He received his MS and PhD degrees in Industrial Engineering from the University of Pittsburgh. His dissertation received the Pristker Doctoral Dissertation Award from Institute of Industrial and Systems Engineering in 2013. His research interests span theoretical, computational, and applied aspects of mathematical programming, focusing on data-driven decision making problems in personalized medicine, public health and healthcare operations management. He is an Associate Editor for the IISE Transactions in the Operations Engineering and Analytics Focused Issue. He currently serves as a council member of the INFORMS Health Applications Society.