

FCSM

# RESEARCH CONFERENCE

**December 1-3, 2015**

**Washington, DC**

# Abstract Submission Template

The conference provides a forum for experts from around the world to discuss and exchange current research and methodological knowledge relevant to statistical programs. **Papers must be original and not previously published or disseminated.**

**Deadline:** Submit the abstract via email to [**fcsm.conf@gmail.com**](mailto:fcsm.conf@gmail.com) by **May 5, 2015**. Late submissions will not be accepted.

Please note that authors/ presenters are responsible for their travel expenses, but the registration fee is waived for one presenter per paper.

**1. Author Contact Information**

**Primary Author Information**

Please provide the following information for the primary author:

|  |  |
| --- | --- |
| **First Name** | Robert |
| **Last name** | Yerex |
| **Organization** | UVA Medical System / Accountable Care Organization |
| **Mailing address** | BOX 800777  Charlottesville, VA 22908 |
| **Country** | USA |
| **Phone** | 503-545-5874 |
| **Email** | ry5t@virginia.edu |
| **Fax** |  |

**Secondary Author(s) Information**

Please provide the full name, affiliation, and email for each co-author, if applicable. Add additional rows if necessary.

|  |  |
| --- | --- |
| **Name and Affiliation** | **Email** |
|  |  |
|  |  |
|  |  |
|  |  |

**2. Topic Area:**

### (Select from list: Survey design and data collection, Evaluation, Cross-cutting topics, Estimation and Analysis, Other – please specify)

|  |
| --- |
| Estimation and Analysis |

**3. Keywords (enter up to 5)**

|  |
| --- |
| **Predictive modeling, feature engineering, CMS Claims Data, Readmission** |
|  |
|  |

**4. Briefly describe the research hypothesis or problem your presentation addresses (limit 120 words)**

|  |
| --- |
| Claims based medical records such as that provided by CMS for the patient members of ACOs, represent a large and complex source of data for the development of predictive models targeting patient outcomes; however the complexity and lack of standardization of claims data is a major impediment to its application. Feature engineering can be used to reduce this data complexity and allow it to be used in the training and validation of effective predictive outcomes. |

**5. What methods and/or data are used to answer the question or solve the problem? (limit 120 words)**

|  |
| --- |
| Using only data provided by the CMS, statistical learning technique, in the form of a random survival forest algorithm, was used to develop a model for dynamically predicting the hazard of unplanned readmission during the critical first 30 days post discharge. This model included several unique engineered features. |

**6. Briefly describe what the audience will take away from your presentation (how is it applicable to other researchers?). (limit 120 words)**

|  |
| --- |
| Exposure to the application of statistical learning techniques for the development of patient outcome targeted predictive models and the need for sophisticated feature engineering associated with the development of these models. |

**7. Please indicate the submission type:**

\_X\_\_ Contributed paper

\_\_\_\_ Organized session paper\* (please list other papers included in the proposed session below)

\_\_\_\_ Technical software demonstration\*\*

**\*Organized sessions**: Proposals for organized sessions should include 3 papers and a discussant OR 4 papers around a topical area but should represent more than one agency and/or program. Complete the template for each paper in the proposed session. The session organizer should email the abstract files together in one email to FCSM. Also, include the session title, session description, and discussant name (if applicable) in your submission.

**\*\*Technical software demonstration:** There will be an opportunity to present brief demonstrations of new software or data access tools. Note that these should be informational demonstrations and not marketing displays.

**8. Abstract (limit 300 words):**

Please use the following formatting instructions when preparing your abstract:

Font type: Arial 12pt.

Font style Title – bold type

Text – regular type

Case: Upper and lower-case letters for all headings and text

Line spacing: Single

Paragraph format: No initial indent, use blank line between paragraphs

Authors: Please list each author’s name under the abstract title followed by their affiliation in parentheses

See the next page for an example.

**This is the Abstract Title**

Robert P. Yerex, UVA Medical System

The timing of post-discharge care is a significant factor in reducing unplanned hospital readmission. Statistical learning techniques can be applied to the development of models that predict the likelihood of patient readmission during the critical 30 day post-discharge period. The accuracy of these models is dependent on the quantity and quality of data used for training and validation. For Medicare and Medicaid patients who are members of an Accountable Care Organization (ACO), the Centers for Medicare and Medicaid (CMS) provides detailed claims based data that can be used, if appropriately collated, and transformed. This involves the identification and creation of useful features, which when included in the model, increases its predictive strength. Creation of derived features (feature engineering) is a process in which a large number of base dimensions (n) are combined to create a smaller features set (n\* << n), reducing the complexity of the model, while retaining its inherent information value. The inpatient admission diagnoses, in the form of ICD9 codes, are an example of high dimensionality attributes found in the CMS claims data. Patterns inherent in the combinations of these codes can be used to create an engineered feature. In this study, a taxonomy of patterns of patient diagnoses was developed that was then used as a feature within a random survival forest model that predicts the hazard function (where the hazard event is unplanned readmission) of an individual patient for the first 30 days post discharge. Over the ensuing 30 days after release from hospital, a patient’s likelihood of readmission can be dynamically estimated based on the remaining portion of the hazard curve. Inclusion of the multiple diagnoses feature increased model accuracy to the point where it could be effectively used as a tool for targeting post-discharge patient care.