1493. Creation and External Validation of a Clinical Prediction Rule for Diarrheal Etiology Using Natural Language Processing
Joel Iverson, Howard, MD; Joni Aoki, Jeff Ferraro, PhD, Ben Haaland, PhD; Andrew Pavia, MD, Daniel T. Leung, MD, MSc; University of Utah, Salt Lake City, Utah
Session: 158. Enteric and Intraabdominal Infections
Friday, October 4, 2019: 12:15 PM
Background. Infectious diarrheal illness is a significant contributor to health-care costs in the US pediatric population. New multi-pathogen PCR-based panels have shown increased sensitivity over previous methods; however, they are costly and clinical utility may be limited in many cases. Clinical Prediction Rules (CPRs) may help optimize the appropriate use of these tests. Furthermore, Natural Language Processing (NLP) is an emerging tool to extract clinical history for decision support. Here, we examine NLP for the validation of a CPR for pediatric diarrhea.

Methods. Using data from a prospective clinical trial at 5 US pediatric hospitals, 961 diarrheal cases were assessed for etiology and relevant clinical variables. Of 65 variables collected in that study, 42 were excluded in our models based on a scrutiny of documentation in reviewed clinical charts. The remaining 23 variables were ranked by random forest (RF) variable importance and utilized in both an RF and stepwise logistic regression (LR) model for viral-only etiology. We investigated whether NLP could accurately extract data from clinical notes comparable to study questionnaires. We used the eHOST abstraction software to abstract 6 clinical variables from patient charts that were useful in our CPR. These data will be used to train an NLP algorithm to extract the same variables from additional charts, and be combined with data from 2 other variables coded in the EMR to externally validate our model.

Results. Both RF and LR models achieved cross-validated area under the receiver operating characteristic curves of 0.74 using the top 5 variables (season, age, bloody diarrhea, vomiting/nausea, and fever), which did not improve significantly with the addition of more variables. Of 270 charts abstracted for NLP training, there were 41 occurrences of bloody diarrhea annotated, 339 occurrences of vomiting, and 145 occurrences of fever. Inter-annotator agreement over 9 training sets ranged between 0.63 and 0.83.

Conclusion. We have constructed a parsimonious CPR involving only 5 inputs for the prediction of a viral-only etiology for pediatric diarrheal illness using prospectively collected data. With the training of an NLP algorithm for automated chart abstraction we will validate the CPR. NLP could allow a CPR to run without manual data entry to improve care.

### Table 3: Management and outcomes of the patients

| Characteristics | Cancer | Non cancer | P-value |
|----------------|--------|------------|---------|
| Biliary drainage | 87 (86.1) | 24 (43.6) | <0.0001 |
| Endoscopic | 64/87 (73.6) | 21/42 (87.5) | 0.18 |
| Radiologic | 3/87 (4.25) | 3/24 (12.5) | 0.008 |
| More than one session | 25/87 (29.2) | 2/24 (8.3) | <0.0001 |
| Drainage timing, days | 3.6 (2.5) | 3.4 (2.4) | 0.87 |
| Antibiotic treatment duration, days (mean ± SD) | 13.6 (9.9) | 11.3 (8.6) | 0.17 |

Outcome

| Favourable outcome* | 52 (51.5) | 47 (85.5) | <0.0001 |
| Death at day 27 following acute cholangitis | 22 (23.3) | 1 (1.8) | 0.0007 |
| Fever relapse before 28 days | 17 (16.8) | 3 (5.5) | 0.047 |
| Liver abscess | 5 (5.0) | 1 (1.8) | 0.43 |

SD: Standard Deviation
* patient alive, andfebrile and with no liver abscess and no need for a secondary drainage at day 28.

### Disclosures. All authors: No reported disclosures.