# Using natural language processing to optimize case ascertainment of acute otitis media in a large, state-wide pediatric practice network

Joshua C Herigon MD, MPH, MBI[1], Amir Kimia, MD[1], Marvin Harper, MD[1]
[1]Boston Children's Hospital & Harvard Medical School, Boston, MA

Boston Children's Hospital

## Background

- To help address inappropriate antibiotic prescribing, the CDC has included "audit with feedback" as a core component of outpatient antibiotic stewardship
- Previous research employing "audit with feedback" in outpatient settings relied on diagnosis codes provided by prescribers to categorize antibiotic use and appropriateness
- Diagnosis codes may be inaccurate due to:
  - Human error
  - Lack of awareness of proper coding
  - Diagnosis shifting to justify a prescription
- Natural language processing can analyze free text to extract specific semantic concepts or classify text
- Using natural language processing, we sought to identify cases of acute otitis media (AOM) based on clinical documentation

## Methods

**Study Design:** cross-sectional retrospective chart review

**Setting/Population**

- 80+ independently-owned pediatric practices affiliated with Boston Children's Hospital, includes 400+ clinicians taking care of > 400,000 children
- Patients < 5 years old
- Encounters July 1, 2018 - June 30,2019
- Problem-focused, in-person visits only
- 12 randomly selected weekdays (one/month) plus an additional random weekday in Jan 2019 for validation
- Complete note text and limited structured data extracted

### Manual Labeling

- Key physical exam descriptors (see table) defined based on the AAP AOM guideline
- Notes were human reviewed and manually labeled as "AOM present" or "AOM absent" based on the presence or absence of these descriptors

| Term | Examples of acceptable synonyms |
|---|---|
| bulging | bulge, pus filled, swelling, distorted landmarks, obscured landmarks, full |
| purulence | pus, opaque, turbid, yellow or white fluid, thick, thick fluid, straw colored, mucoid |
| erythema[a] | red, injected, inflammed, pink, erythematous, vessels visible |
| effusion[b] | dull, no light reflex, abnormal light reflex, serous, serous fluid, clear fluid, fluid level |
| otorrhea[c] | drainage, TM ruptured, discharge, fluid in the canal |

[a]Must be combined with an indication for bulging or purulence
[b]Must be combined with an indication for erythema
[c]Not due to otitis externa

**Corresponding Author**
Joshua Herigon MD, MPH, MBI
joshua.herigon@childrens.harvard.edu
www.joshherigon.com

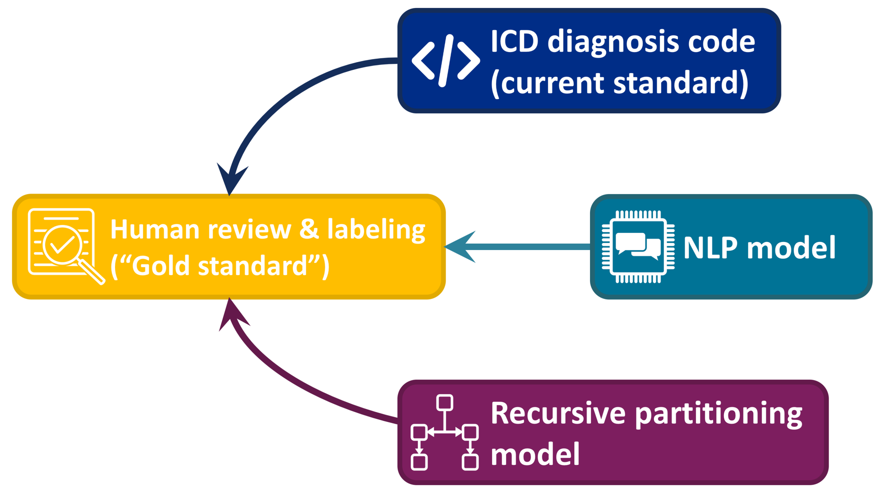
**Acknowledgements**
Joshua Herigon was supported by a National Library of Medicine Biomedical Informatics and Data Science Research Training Grant (T15LM007092)

### Natural Language Processing (NLP)

- A supervised machine learning model (support vector machine or SVM) using n-grams trained to automatically identify positive and negative instances
- SVM produces a score, with higher scores more likely to be positive
- Two different cutoffs of SVM values were employed
  - First cutoff balanced sensitivity and specificity to optimize the NLP model alone
  - Second cutoff chosen automatically by multivariate recursive partitioning model

### Recursive partitioning (RP) model

- Form of multivariate analysis employing decision trees
- The RP model was created by combining NLP results optimized for specificity with structured data
- Multiple candidate RP models were then created and tuned using the training cohort data to optimize sensitivity before a final model was chosen




### Statistical Analysis

- Human review considered the "gold standard"
- 3 methods compared against human review
- Sensitivity, specificity, positive and negative predictive value (with 95% confidence intervals) was calculated for each

## Results

**Fig 1. Details of encounters included in the training and validation cohorts**

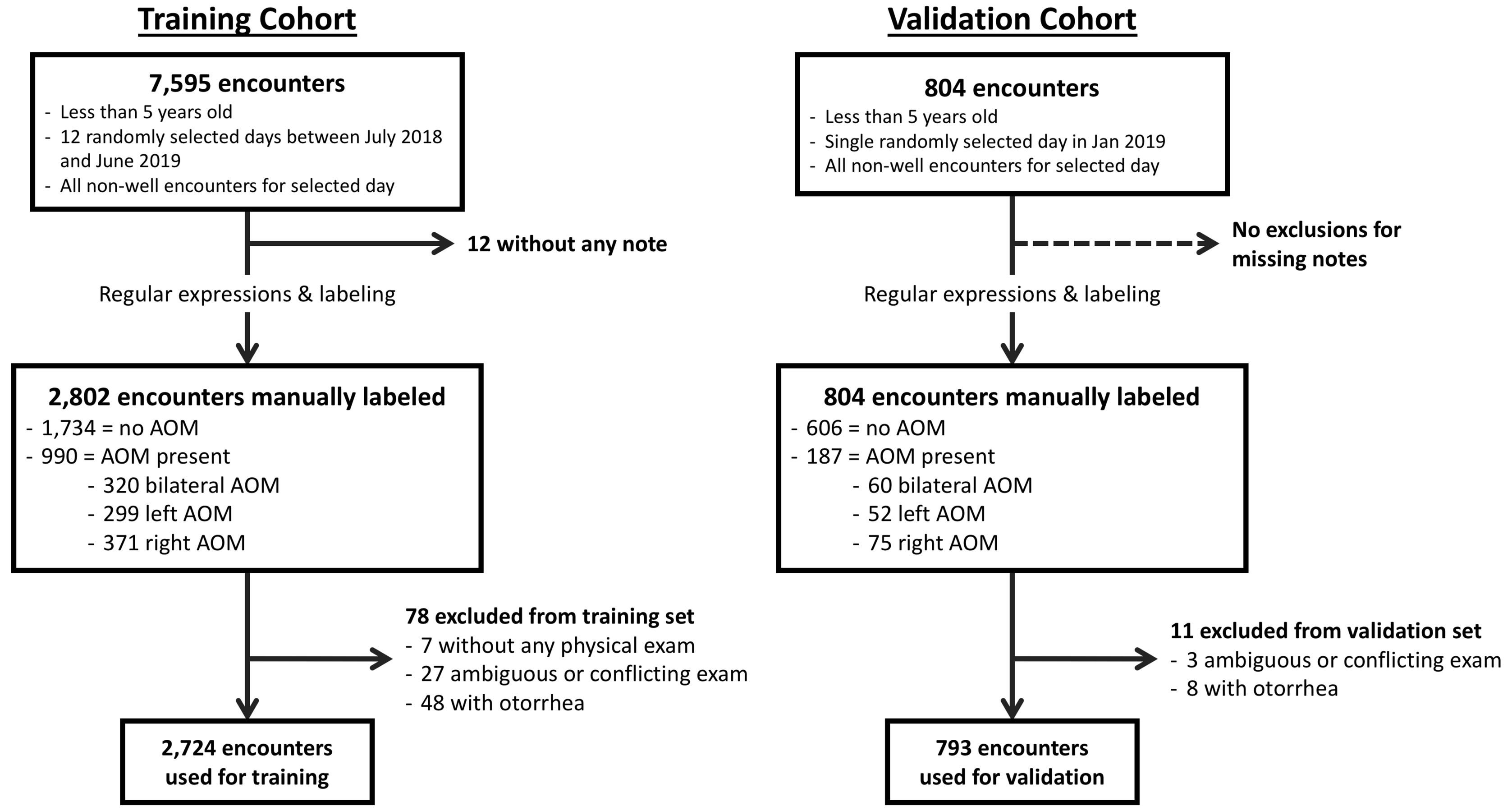



**Table 1. Cohort characteristics**

| | Training (*N* = 2,724) | Validation (*N* = 793) | All Encounters (*N* = 3,517) | *P*-value |
|---|---|---|---|---|
| **Patient characteristics** | | | | |
| Age, median months (IQR[a]) | 25 (14, 40) | 26 (15, 42) | 25 (14, 40.5) | 0.056 |
| Female gender (%) | 1,213 (44.5) | 348 (43.9) | 1,561 (44.4) | 0.747 |
| Complex chronic conditions (%) | 270 (9.9) | 84 (10.6) | 354 (10.1) | 0.575 |
| **Visit characteristics** | | | | |
| Chief complaint (%)[b] | | | | |
| Fever | 471 (17.3) | 165 (20.8) | 636 (18.1) | <0.001 |
| Cough | 454 (16.7) | 180 (22.7) | 634 (18.0) | |
| Ear problem | 440 (16.2) | 101 (12.7) | 541 (15.4) | |
| Follow-up | 279 (10.2) | 68 (8.6) | 347 (9.9) | |
| Rash | 185 (6.8) | 37 (4.7) | 222 (6.3) | |
| AOM diagnosis (%) | 1,068 (39.2) | 194 (24.5) | 1,262 (35.9) | <0.001 |

**Fig 2. Details of the decision tree produced by the recursive partitioning model with performance at each decision node using the training cohort**

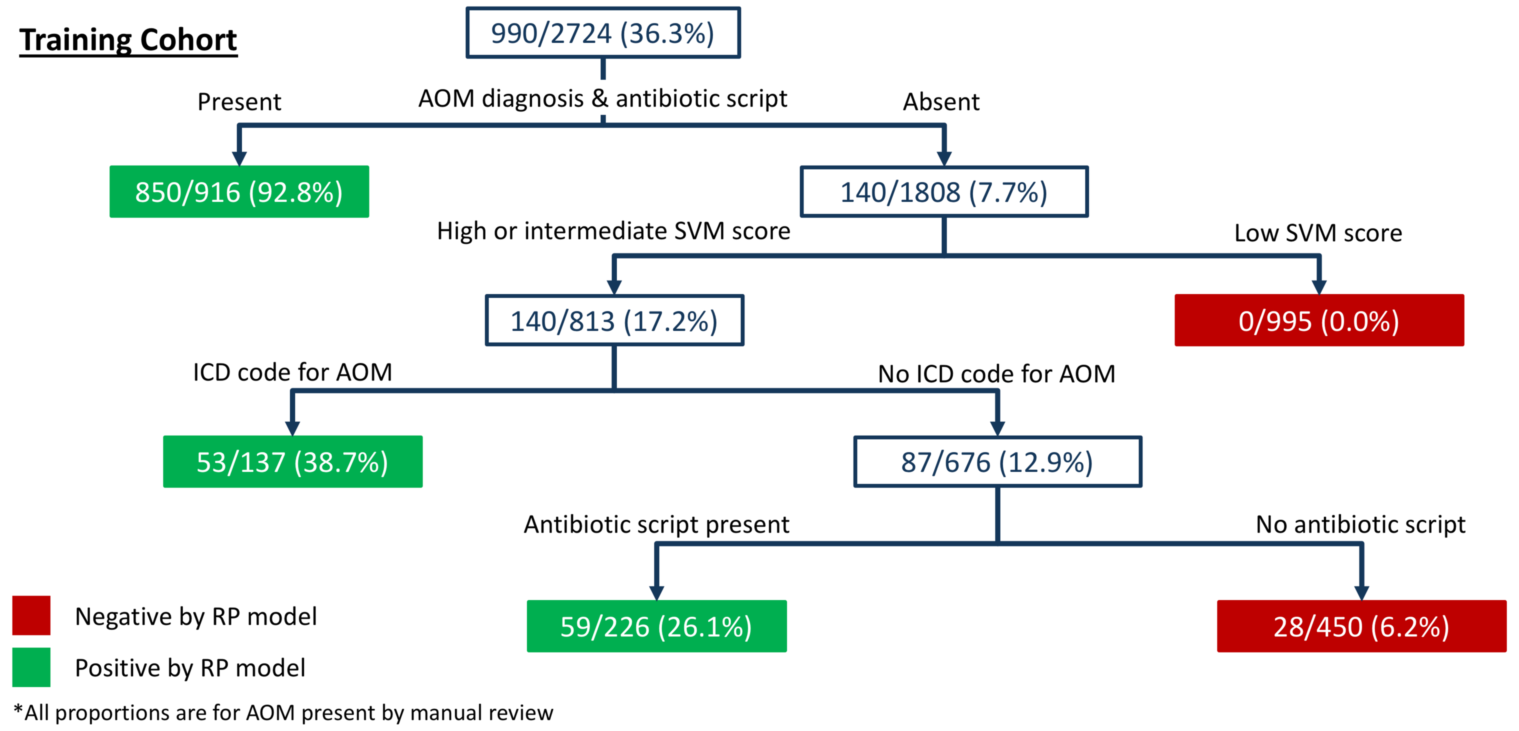



**Table 2. Performance comparison of all 3 methods**

| | | Sensitivity | | | Positive Predictive Value | | | Specificity | | | Negative Predictive Value | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Estimate | % | (95% CI) | Estimate | % | (95% CI) | Estimate | % | (95% CI) | Estimate | % | (95% CI) |
| **Training cohort (N = 2,724)** | ICD codes | 903/990 | **91.2** | (89.2, 92.9) | 903/1068 | **84.6** | (82.2, 86.6) | 1569/1734 | **90.5** | (89.0, 91.8) | 1569/1656 | **94.7** | (93.5, 95.7) |
| | NLP | 922/990 | **93.1** | (91.3, 94.6) | 922/1035 | **89.1** | (87.0, 90.9) | 1621/1734 | **93.5** | (92.2, 94.6) | 1621/1689 | **96.0** | (94.9, 96.8) |
| | RP model | 962/990 | **97.2** | (95.9, 98.1) | 962/1279 | **75.2** | (72.7, 77.5) | 1417/1734 | **81.7** | (79.8, 83.5) | 1417/1445 | **98.1** | (97.2, 98.7) |
| **Validation cohort (N = 793)** | ICD codes | 165/187 | **88.2** | (82.5, 92.3) | 165/194 | **85.0** | (79.1, 89.6) | 577/606 | **95.2** | (93.1, 96.7) | 577/599 | **96.3** | (94.4, 97.6) |
| | NLP | 156/187 | **83.4** | (77.1, 88.3) | 156/184 | **84.8** | (78.6, 89.5) | 578/606 | **95.4** | (93.3, 96.9) | 578/609 | **94.9** | (92.3, 96.5) |
| | RP model | 176/187 | **94.1** | (89.4, 96.9) | 176/236 | **74.6** | (68.4, 79.9) | 546/606 | **90.1** | (87.4, 92.3) | 546/557 | **98.0** | (96.4, 99.0) |

## Conclusions

Natural language processing of outpatient pediatric visit documentation can be used successfully to create models accurately identifying cases of AOM based on clinical documentation. Combining NLP and structured data improves automated case detection. These techniques may be valuable in optimizing outpatient antimicrobial stewardship efforts.