Prospective Validation of the MET-AP Clinical Decision Support System for Pediatric ED Triage of Acute Abdominal Pain
Ken Farion MD
Wojtek Michalowski PhD  Steven Rubin MD
Szymon Wilk PhD  Rhonda Correll RN

Departments of Pediatrics and Surgery
Children’s Hospital of Eastern Ontario
and
The School of Management
University of Ottawa

Mobile Emergency Triage (MET) Study Group

Funded by
Natural Sciences and Engineering Research Council of Canada (NSERC)  
and
Physicians’ Services Incorporated Foundation (PSI) of Ontario
Background

- Abdominal pain is a common complaint
- ED disposition decisions can be difficult and time-consuming
  - large list of possible diagnoses
  - difficulty assessing children and recognizing patterns of presentation
  - lack of simple diagnostic tests
  - lack of valid and easily applied CDR
- Improved efficiency possible with correct path early on
What is MET?

• Mobile Emergency Triage-Abdominal Pain
• Clinical Decision Support System
  ▪ based on data mining of 700 retrospective patients
  ▪ complex algorithm considers as many as 13 historical, physical exam or lab attributes
  ▪ easily determined during initial assessment
  ▪ generates an estimate of how closely the patient matches each of the 3 categories
**What is MET-AP?**

- View triage as the disposition decisions made by the ED physician
- Assists physicians by suggesting which direction to take:
  - consult General Surgery for probable appendicitis
  - discharge, F/U as needed for benign/resolving causes
  - observe/investigate for other causes requiring treatment/management
What is MET?

• Operates on a PDA with simple interfaces
  ▪ functions independently on weak platforms
  ▪ weak connectivity with powerful server

ANYTIME and ANYWHERE

• MET-AP is not a diagnostic test, nor does it replace U/S or other tests
Data Entry of Attributes
Triage Recommendation

Patient: Peterson, James

Suggested: Observation (strong)

Discharge: weak
Observation: strong
Consult: moderate

Main
Objectives

- To determine the accuracy of MET-AP in recommending the correct triage category
- To determine the accuracy of staff/resident in predicting the correct triage category
- To determine the inter-observer agreements between staff and residents for evaluating patient attributes
Methods

• Prospective ED cohort study 24/7
• Patients (1-16y)
• Acute abdominal pain (<10d)
• Exclusion criteria
  • insignificant abdominal pain found
  • trauma
  • prior abdominal surgery
  • symptoms related to a current diagnosis
  • chronic disease
  • direct referral to General Surgery
  • prior enrollment within 4 weeks
  • unable to complete follow-up
  • refused consent
Methods

- Observers completed enrollment and data entry on PDA
- Included as many attributes as deemed necessary
- Entered a prediction of the most appropriate triage category
- Where possible, 2 independent observations & predictions were completed by resident/staff
**Methods**

- Telephone follow-up & chart review was conducted 14 days after the ED visit
- Determine final diagnosis and the corresponding “Gold Standard” triage category
- All personnel blinded to MET-AP triage recommendation throughout
Methods

• MET recommendation defined as the MET triage category with the highest strength

• MET and physician accuracies presented as percentage of times they matched the “Gold Standard” category (95% CI)

• Inter-observer agreements for each attribute
  ▪ Kappa statistic (categorical)
  ▪ Intra-class correlation coefficients (continuous)
Results - Enrollment

Patient Enrollment
July 2003 – Feb 2004

34,527 Visits

2,255 Visits with AP

1,157 Visits Excluded

1,098 Eligible Visits

467 Not Approached

631 Approached

38 Refused Consent

593 Enrolled

12 Excluded Post

574 Analyzed

7 Lost to Follow-up
# Results – Patient Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Approached (n=631)</th>
<th>Not Approached (n=467)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Age (SD)</td>
<td>9.23 (4.01)</td>
<td>9.48 (4.42)</td>
<td>0.349 *</td>
</tr>
<tr>
<td>% Male</td>
<td>48.7%</td>
<td>47.6%</td>
<td>0.760 +</td>
</tr>
<tr>
<td>Triage Category Corresponding to the Final Diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discharge</td>
<td>76.4%</td>
<td>71.9%</td>
<td>0.238 #</td>
</tr>
<tr>
<td>Observe/Investigate</td>
<td>13.1%</td>
<td>15.2%</td>
<td></td>
</tr>
<tr>
<td>Consult</td>
<td>10.5%</td>
<td>12.9%</td>
<td></td>
</tr>
</tbody>
</table>

* Student t-test
* Fisher’s exact test
# Chi-square test
Results - Observers

574 Patients Analyzed

235 Staff only

50 Staff then Resident

172 Resident then Staff

117 Resident only

457 Staff Observations

222 Both Observations

339 Resident Observations
<table>
<thead>
<tr>
<th>Percentage of times correct</th>
<th>Staff Physician (n=457)</th>
<th>Resident (n=339)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MET Recommendation</td>
<td>67.2% (95% CI: 62.7, 71.3)</td>
<td>63.4% (95% CI: 58.2, 68.4)</td>
</tr>
<tr>
<td>Physician Prediction</td>
<td>70.2% (95% CI: 65.9, 74.2)</td>
<td>62.8% (95% CI: 57.6, 67.8)</td>
</tr>
<tr>
<td>Difference</td>
<td>-3.0%</td>
<td>0.6%</td>
</tr>
</tbody>
</table>
# Results - MET & Staff Physician vs Final Diagnosis Category

<table>
<thead>
<tr>
<th>Final Outcome Category</th>
<th>MET Recommendation</th>
<th>Physician Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disc</td>
<td>Obs/Inv</td>
</tr>
<tr>
<td>Benign</td>
<td>262</td>
<td>67</td>
</tr>
<tr>
<td>Other</td>
<td>39</td>
<td>11</td>
</tr>
<tr>
<td>Appendicitis</td>
<td>12</td>
<td>2</td>
</tr>
</tbody>
</table>
### Inter-observer Agreements

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>(95% CI)</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration (hours)</td>
<td>0.961 (0.949, 0.970)</td>
<td>ICC</td>
</tr>
<tr>
<td>Duration (categorized)</td>
<td>0.825 (0.751, 0.900)</td>
<td>K</td>
</tr>
<tr>
<td>Vomiting</td>
<td>0.890 (0.829, 0.951)</td>
<td>K</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>0.970 (0.961, 0.977)</td>
<td>ICC</td>
</tr>
<tr>
<td>Temperature (categorized)</td>
<td>0.945 (0.902, 0.989)</td>
<td>K</td>
</tr>
<tr>
<td>Total WBC</td>
<td>0.942 (0.891, 0.970)</td>
<td>ICC</td>
</tr>
<tr>
<td>WBC (categorized)</td>
<td>0.952 (0.860, 1.000)</td>
<td>K</td>
</tr>
</tbody>
</table>

*K* = Kappa statistic for categorical data

*ICC* = Intra-class correlation coefficient for continuous data
## Inter-observer Agreements

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Inter-observer Agreement (95% CI)</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site of Pain</td>
<td>0.513 (0.409, 0.617)</td>
<td>K</td>
</tr>
<tr>
<td>Intermittent vs Constant</td>
<td>0.475 (0.356, 0.594)</td>
<td>K</td>
</tr>
<tr>
<td>Shifting of Pain</td>
<td>0.521 (0.362, 0.680)</td>
<td>K</td>
</tr>
<tr>
<td>Previous ER Visit</td>
<td>0.481 (0.128, 0.834)</td>
<td>K</td>
</tr>
<tr>
<td>Site of Tenderness</td>
<td>0.573 (0.467, 0.679)</td>
<td>K</td>
</tr>
<tr>
<td>Rebound Tenderness</td>
<td>0.449 (0.266, 0.632)</td>
<td>K</td>
</tr>
<tr>
<td>Localized Guarding</td>
<td>0.309 (0.139, 0.479)</td>
<td>K</td>
</tr>
</tbody>
</table>

*K = Kappa statistic for categorical data

*ICC = Intra-class correlation coefficient for continuous data
Discussion

- Considered only the MET triage category with the highest strength
- The system provides a strength for all 3 categories, mimicking the way physicians think through the problem
- How often was the correct triage category recommended as a close second-place?
Discussion

• MET-AP has poor accuracy for the *Observation* category, likely due to the diverse range of cases included, while physicians had additional data to make their prediction.

• MET-AP recommendations tend to be skewed towards *Discharge*, likely due to the relative prevalence of Benign presentations.
Discussion

• Inter-observer agreement between staff physicians and residents was poorest for subjective attributes like rebound tenderness and involuntary guarding.

• Does this lower agreement affect the system’s triage accuracy?

• Which observer should be considered the gold standard?

• What is the agreement between two observers at the same level?
Discussion

• The quality and completeness of data collected during the trial is far superior to the retrospective data used to develop MET-AP

• A new algorithm based on this prospective data should yield improved accuracy
Conclusions

• MET-AP recommends the correct triage category with similar accuracy to ED staff and resident physicians

• Most attributes have moderate to near-perfect inter-observer agreement
Further Information on MET

www.mobiledss.uottawa.ca

- Background on MET research
- Other MET modules
  - hip pain
  - scrotal pain
  - asthma exacerbation
- MET publications
- Downloadable version of MET-AP