Prospective Evaluation of the MET-AP System Providing Triage Plans for Acute Pediatric Abdominal Pain

Ken J. Farion[†], MD, FRCPC
Departments of Pediatrics and Emergency Medicine, University of Ottawa
Ottawa, Canada

Address for correspondence:
Emergency Department, Children's Hospital of Eastern Ontario
401 Smyth Road
Ottawa, ON Canada, K1H 8L1
tel.: (613) 737-7600 x. 2899, fax: (613) 738-4885
e-mail: farion@cheo.on.ca

Wojtek Michalowski, PhD
MET Research Group, University of Ottawa
Ottawa, Canada

Steven Rubin, MD, FRCSC, FRCSE
Department of Surgery, University of Ottawa
Ottawa, Canada

Szymon Wilk[‡], PhD
Institute of Computing Science, Poznan University of Technology
Poznan, Poland

Rhonda Correll, BScN (Hon)
Children's Hospital of Eastern Ontario
Ottawa, Canada

Isabelle Gaboury, PhD(C)
Chalmers Research Group
Ottawa, Canada

Keywords: pediatric abdominal pain; hospital, triage; decision support system, clinical; computer, handheld

[†] Corresponding author
[‡] This research was conducted when Szymon Wilk was a post-doctoral fellow at the MET Research Group, University of Ottawa

ABSTRACT

Background: Children with acute abdominal pain (AP) are frequently assessed in the Emergency Department (ED). Though the majority of patients have benign causes, uncertainty during the physician's initial assessment may result in unnecessary tests and prolonged observation before a definitive disposition decision can be made. A rule-based mobile clinical decision support system, MET-AP (Mobile Emergency Triage – Abdominal Pain), has been developed to recommend an appropriate triage plan (discharge, consult surgery, or observe/investigate) early in the ED visit, with the goal of promoting ED efficiencies and improved patient outcomes.

Objective: To prospectively evaluate the accuracy of MET-AP to recommend the correct triage plan when used during the initial assessment by staff emergency physicians (EPs) and residents in a tertiary care pediatric ED.

Design: Prospective cohort study. Staff EPs and/or residents examined children, aged 1 to 16 years, with acute, non-traumatic AP of less than 10 days duration. Details of their initial assessment, along with their blinded prediction of the correct triage plan, were recorded electronically. Inter-observer assessments were collected, where possible. Telephone and chart follow-up at 10 to 14 days was conducted to determine the patient's outcome/diagnosis, and thus the gold standard triage plan appropriate for the patient's visit.

Measurements: Accuracy of MET-AP to recommend the correct triage plan (i.e., to match the gold standard plan); accuracy of physicians to predict the correct triage plan; inter-observer agreement between staff EPs and residents for each clinical attribute recorded within MET-AP.

Results: Over eight months, 574 patients with AP completed follow-up (10% appendicitis, 13% other pathology, 77% benign/resolving conditions). For patient assessments by the staff EP (n=457), the MET-AP recommendation was correct for 72% of patients (95% CI's: 67.9 to 76.1), while the physician's prediction was correct in 70% of cases (65.9 to 74.2)(p=0.518). However, staff EP triage plans were more conservative than those generated by MET-AP, and a small number of patients whose triage plan should have been "consult surgery" would have been "discharged" by MET-AP. For resident assessments (n=339), MET-AP and physician accuracies were slightly lower, but not statistically different from staff results or from each other. Inter-observer agreement on most attributes was moderate to near perfect.

Conclusion: MET-AP shows promise in recommending the correct triage plan with similar overall accuracy to experienced pediatric EPs, but requires further research to improve accuracy and safety. MET-AP can be used on all pediatric ED patients with AP and is capable of producing a triage plan recommendation without requiring a complete set of patient information.

What is already known on this topic
Abdominal pain is a common complaint in children, and it can be difficult to differentiate between serious causes (appendicitis) and more common benign causes (constipation, viral illnesses). Attempts to develop abdominal pain scores have resulted in modest success identifying appendicitis in small pre-selected subsets of the presenting population, but add little to the assessment of most children with abdominal pain presenting to the ED. Patient-specific clinical decision support systems have great potential. However, there is limited knowledge of how these systems perform prospectively.
What this study adds
This is the first comprehensive prospective evaluation of a clinical decision support system that recommends patient-specific ED triage plans on the correct initial course of management for a diverse group of children with acute abdominal pain. MET-AP differs from traditional appendicitis scores by recommending a triage plan without the strict requirement of having a complete set of patient information.

1. INTRODUCTION

Acute abdominal pain (AP) in children is a common but challenging problem for emergency department (ED) personnel, representing 6.5% of ED visits at our tertiary-care pediatric institution (2004-2005 data). While some patients have serious conditions requiring admission or surgery, most have benign causes. Symptoms frequently resolve without complication before a cause is found so that a definitive diagnosis is not possible during the ED visit. In these cases, choosing the correct triage plan in an efficient manner is an important proxy. In this research, we define the triage plan as an ED activity that extends beyond the initial assessment and categorization typically completed by a triage nurse to include the initial assessment and management decisions made by the emergency physician (EP). The triage plan may result in discharging the patient with non-serious complaints or may involve diagnostic tests, specialty consultations, and interim management, which may lead to a definitive diagnosis.

Developing a triage plan is complicated by the inherent uncertainty about conditions causing AP [1]. Experienced EP's offset this uncertainty with their clinical expertise to combine key historical information and physical findings occurring in recognizable patterns to choose the correct triage plan quickly and efficiently [2]. Inexperienced physicians may have weaker information-gathering and decision-making skills [2-4], resulting in reliance on expensive and time-consuming tests while continuing to be uncertain about the cause or seriousness of the patient's complaints.

The majority of the literature dealing with ED management of AP has focused on identifying patients with appendicitis, the most common surgical emergency of childhood [5], as opposed to the broader triage plan that is the subject of our research. While the diagnosis of appendicitis can be assisted in some cases by the use of diagnostic imaging [6-9] or laboratory tests [10-12], these investigations are time-consuming, expensive, and not always readily available. A number of appendicitis scores [13-15] have been developed by identifying the most differentiating clinical symptoms and signs that, in totality, will produce a given score. To obtain a result, therefore, the physician is forced to gather all required patient information on every patient, often including a white blood cell (WBC) count. However, these scores do not address what to do with patients without appendicitis, who are the majority of patients with AP in the ED. Finally, prospective validation of these scores, even when applied only to patients suspected of having appendicitis to mimic the populations they were initially developed on, have failed to produce acceptable operating characteristics to allow these scores to be routinely incorporated into patient management [16-20].

Clinical decision support systems (CDSS) are computer systems that directly or indirectly assist healthcare professionals in making clinical decisions [21]. Systems for managing information (e.g., electronic patient records) and for focusing attention to critical information (e.g., laboratory monitors) have been used in clinical practice for a long time and are well accepted [22]. Interest in systems that provide patient-specific recommendations is increasing, in parallel with demands to improve healthcare processes and patient outcomes in a cost-efficient manner [23]. Despite many significant advances, most CDSS have yet to undergo comprehensive evaluation to determine their impact on patient outcomes or healthcare systems beyond limited trial environments [24], and there exists a significant void of knowledge on factors influencing adoption and integration of these technologies into existing clinical workflows [25].

Attempts to develop computer-based systems providing patient-specific recommendations for acute AP date back thirty years, beginning with work by de Dombal [26-28] who was primarily interested in supporting and standardizing the clinical decision making process. de Dombal's original system is based on a Bayesian model of verified cases to generate a probabilistic assessment of the differential diagnosis of the patient in question, using the information collected by physicians on structured data sheets. Various decision models (i.e., ED, Ward, Gynecology, and Pediatric) have been developed, each with their own set of possible diagnoses to consider. While the adult models consider multiple diagnoses, the pediatric model only considers appendicitis and non-specific AP. Despite early promising results, no subsequent system has been widely adopted nor found through appropriate evaluation to significantly impact clinical decision making [29]. Improvements in the diagnostic accuracy of physicians using the system have been attributed, not to the CDSS, but to the structured data collection form that stimulated less experienced physicians to obtain and consider a complete and organized history and physical exam [30-32]. Assessment of the system's ability to correctly diagnose or suggest appropriate management has been limited [29].

Another example of a patient-specific CDSS for AP is MEDUSA [1]. This hybrid expert system combined rule-based reasoning, heuristic reasoning and case-based reasoning to process data on up to 200 symptoms to generate a list of probable and possible diagnoses. A limited retrospective evaluation of 100 cases has yielded an accuracy of 80%; no prospective evaluation has been reported. While an accuracy of 80% is encouraging, such a system would not likely be adopted in an ED setting as it may require up to 10 minutes to input as many as 200 symptoms for each patient assessment.

Finally, other researchers have developed a Bayesian model to provide decision support to the triage decisions made by nurses in person or by telephone, potentially acting as gate-keepers to seeking ED care [33]. The AP model of this system underwent a limited retrospective assessment, evaluating the system's 4-point categorization compared to an ED physician's evaluation of abstracted data from 90 patients, and comparing these predictions to the gold-standard of the patient's outcome. Overall, both the triage system and the physician were correct 56% of the time.

We have taken a different approach to decision support for AP management. Instead of focusing on the final AP diagnosis, we chose to provide support early in the ED process by recommending appropriate triage plans. Moreover, instead of asking the EP to determine and input values for all history, physical, and/or tests attributes, as it is required in most other CDSS and AP scores, we allow for the EP to collect patient specific information that they deem pertinent to the presentation. The decision model then generates a weighting for all possible triage plan recommendations based on how well the patient's data matches each triage plan.

The MET (Mobile Emergency Triage) system implements our approach in the form of a modular CDSS that assists EPs with the assessment and selection of a triage plan for pediatric patients with various acute problems [34, 35]. Recognizing the need to provide support at the point of care in this busy clinical environment, we have deployed MET on mobile devices (e.g., handheld computers) and interfaced it with existing hospital information systems (e.g., registration, laboratory), reducing duplication of data entry.

The AP decision support system [36-39] (MET-AP) discussed in this paper consists of user-friendly structured data entry screens, the clinical algorithm, and a graphical output of how strongly a given patient matches each of the triage recommendations (Figure 1). Retrospective data on 600 patients were used to develop a rule-based decision model using rough set theory [40], augmented with recent extensions to deal with incomplete data [41] and to assess the information value of attributes [42]. Methodologies based on rough sets theory allow an approximation of knowledge from both quantitative and qualitative data, and can handle inconsistencies and incomplete data without removing those records prior to analysis. This was important, as most patient records were incomplete due to the retrospective nature of data collection, and the fact that not all patients required all attributes to be obtained (e.g., not all patients had blood work).

Thirteen attributes (Table 1), found to be most commonly documented and discriminating, were selected. They were used to induce a rule-based model representing the patterns found in the data, selected for its superior predictive power (i.e., ability to accurately triage a new patient) and its acceptable format of clinical knowledge for clinicians. Rules, being “if … then …” logical statements, consisted of several conditions on attribute values, the triage decisions represented by the rule, and the relative strength of the rule, which characterized the generality of the rule (i.e., the higher the relative strength, the more cases from the learning data set are correctly classified by the rule, and the more general pattern the rule represents). A similarity-based classification strategy [43] was used as the solving method to apply the rules to new patient data. The overall strength of each triage recommendation thus was determined as the summation of the strength factors for rules that were satisfied or partially satisfied.

In this way, recommendations can be generated without complete patient data (a very common situation in the ED). The physician can consider the relative strength of each recommendation and combine this advice with their own clinical judgment. We chose this format of decision support because MET-AP is not a diagnostic tool, nor an appendicitis score, but should be viewed as a facilitator (or helper) that focuses the EP’s attention towards the most appropriate plan of care.

The objectives of this prospective evaluation of MET-AP were to: 1) establish MET-AP's overall accuracy to recommend the correct triage plan when used by staff EPs and residents; 2) estimate staff EPs' and residents' overall accuracy to independently predict the correct triage plan; and 3) determine the inter-observer agreement of the clinical attributes collected with MET-AP. The gold-standard categorical triage plan established from the patient's final outcome/diagnosis at follow-up conducted 10-14 days later was used to evaluate each recommendation or prediction.

2. METHODS

2.1 Study Design and Setting

A prospective validation trial of MET-AP was conducted between July 2, 2003 and February 29, 2004 at the Children's Hospital of Eastern Ontario (CHEO) ED, with approval of the CHEO Research Ethics Board. CHEO is a tertiary-care, university-affiliated pediatric hospital with 55,000 annual ED visits. It is the only pediatric ED serving an estimated pediatric population of over 400,000 children in Eastern Ontario and Western Quebec. The ED is staffed with pediatric emergency medicine specialists, augmented by numerous part-time EPs, fellows, and resident trainees. All clinicians except medical students were eligible to participate in a trial; they attended a brief orientation session prior to enrolling patients.

2.2 MET-AP Trial Implementation

For the trial, MET-AP was deployed on several handheld computers (Palm m515; Palm Inc., Santa Clara, CA) that were synchronized with an independent trial server through a wired connection. The server received HL7 data messages from the hospital registration system (EPIC ADT; EPIC Systems Corp., Madison, WI) as patients were registered in the ED. These messages contained basic demographic information and whether the patient was identified as having AP. A dynamic list of current patients with AP (i.e., those registering within the last 8 hours) was stored in the database managed by Microsoft SQL-Server (Version 2000; Microsoft Corp., Redmond, WA).

A trial helper application was added to this deployment of MET-AP to allow physicians to screen patients for eligibility at enrollment using the handheld computer. An additional data entry screen was created to allow the physician to record their independent prediction of the patient's most appropriate triage plan following their initial assessment (Figure 2). Finally, the triage plan support functionality and recommendation output screen was disabled, so that EPs remained blinded to the MET-AP triage plan recommendation. This was necessary for ethical reasons so that the MET-AP triage plan recommendations did not influence the EP's management.

2.3 Study Population

Patients aged 1 to 16 years presenting to the ED with acute AP of less than 10 days duration were eligible for enrollment. All patients had AP as a main presenting complaint or as a significant finding when examined. Patients were excluded for any of the following reasons: 1) insignificant (resolved) AP on history and physical, as determined by the EP; 2) AP following trauma; 3) AP as a complication of a recently diagnosed acute disease process (e.g., urinary tract infection now presenting with pyelonephritis); 4) AP caused by a chronic underlying illness (e.g., inflammatory bowel disease, sickle cell disease, recurrent nephrolithiasis or ovarian cysts); 5) previous intra-abdominal surgery; 6) direct referral to pediatric surgery; 7) prior enrollment within 4 weeks; or 8) inability to complete follow-up (i.e., language barrier or lack of a telephone) as required by study design.

2.4 Patient Recruitment and Enrolment

The treating staff EP or resident approached potentially eligible patients 24 hours per day, 7 days per week, for participation, provided that the physician had been oriented to the trial and MET-AP. Informed written consent was obtained from the patient and/or parent. The physician screened for eligibility and collected all data using the MET-AP system. EPs were instructed to only record data for those attributes they felt were relevant to the patient's presentation; white blood cell (WBC) counts were not obtained if the physician felt this information would not influence his/her management decision. Finally, the physician entered his/her independent prediction of which triage plan the patient was most likely to fit (Figure 2). This prediction was made at the time of initial assessment prior to obtaining an abdominal ultrasound or surgical consult. Beyond this initial assessment, patients were managed in the usual fashion at the discretion of the treating physician, without the influence of a triage plan recommendation from MET-AP.

Where possible, a second observer of the opposite physician level (i.e., staff EP → resident, resident → staff EP) independently assessed the patient within one hour of the primary assessment and recorded their findings and prediction of the triage plan using the MET-AP system, generating a second record for the patient's visit.

2.5 Patient Follow-up

A research assistant (RA) performed a daily audit of all ED visits to identify missed patients with AP. The RA contacted each enrolled patient via telephone 10 to 14 days after their visit for a semi-structured interview, in addition to reviewing the patient's hospital file. This information established a final diagnosis or outcome for each patient's ED visit. From a list of common diagnoses developed by expert opinion and literature review a priori (Table 2), a triage plan appropriate for the patient's visit was established as the gold standard for that visit. This was done blinded to the physicians' prediction and MET-AP triage plan recommendation for the patient. Patients not fitting the list were assigned an appropriate triage plan through consensus of the investigators. Patients not reached after several telephone attempts were sent a letter requesting that they contact the RA to complete the follow-up.

2.6 Study Outcome Measures and Analysis

At the end of the trial, the MET-AP triage plan recommendation for each patient assessment was unmasked. The recommendation consisted of a quantitative strength weighting for each of the three triage plans, corresponding to how closely the patient's data satisfied the rule-based model for each triage plan. The plan with the highest weighting was identified as the final MET-AP recommendation for comparison to the gold standard triage plan determined through patient follow-up. This simplification of the MET-AP recommendations (i.e., selecting one triage plan over the complete advisory information that would normally be provided to the EP about all three triage plans) was done to yield a categorical result that could be compared to the gold-standard triage category. The MET-AP triage plan recommendations or EP triage plan predictions (recorded during the assessment) were considered correct only if they matched the gold standard categorical triage plan established through follow-up.

Patient assessments were grouped according to physician level (i.e., staff EP, resident). For each group, the accuracies of the MET-AP triage plan recommendations and the physician predictions were calculated as the percentage correct with 95% confidence intervals (95% CI) using the Wilson score [44]. Within each physician level group, the accuracy of the two triage plan methods was compared (i.e. MET-AP recommendation versus physician prediction) using sign tests to account for the fact that two predictions were made for each patient. Similarly, within each triage plan method, the prediction accuracy was compared between the physician levels for the subset of patients assessed by both staff EPs and residents.

For each clinical attribute collected, inter-observer agreements were determined using the subset of patients for whom two physician assessments were recorded. Agreement for categorical data was measured using Kappa statistic, while an intra-class correlation coefficient was calculated for continuous data. Both are reported with 95% CI's. Finally, the Kappa statistic was used to assess the agreement between staff EP and resident predictions of the triage plan.

2.7 Sample Size Calculations

Sample sizes calculated a priori used the assumption that physician accuracy predicting the correct category for disposition after initial assessment could be as low as 50%, as reported in the literature. From pilot study data, we had estimated the accuracy of MET-AP to be approximately 80%. Therefore, we determined that MET-AP accuracy and physician prediction accuracy for each physician group could be estimated within ±5%, 19 times out of 20, with a sample of 384 patients. We did not specifically seek to show a difference between physician types or prediction methods. To determine inter-observer agreement between staff and residents for each clinical attribute, we expected that 60% of patients would have assessments recorded by both staff and resident physicians To have 384 patients with paired assessments for analysis, a total of 640 patients would be need to be enrolled.

3. RESULTS

3.1. Description of Patients

During the study period there were 2,255 AP visits to the ED. Figure 3 shows enrolment path of these patients, resulting in a final sample of 574 patients for analysis. Of the 1157 patients excluded, 36% had insignificant or resolved AP, 21% were excluded for age, 20% had prolonged duration of pain, and 10% had underlying chronic illness. No selection bias was found between patients in the final analysis group and those not approached (Table 3), showing that MET-AP was applied to a diverse and representative EP population. The 12 patients initially included then later excluded had various causes for exclusion (e.g., age, prior surgery, duration of pain); no follow-up was completed on these patients. Seven patients were lost to follow-up despite repeated telephone and mail attempts; all were local residents discharged from the ED with benign conditions and none returned to the hospital during the study period. Considering that CHEO is the only facility in the region providing pediatric surgical or inpatient services, we are confident that these children did not have appendicitis or other serious pathology.

Of the 574 patients in the final analysis, 235 patients had only the staff EP record an assessment, 117 patients had only the resident record an assessment, and 222 patients had assessments by both levels of physician (38.7% inter-observer rate). This resulted in 457 staff EP assessments and 339 resident assessments. Over the study period, approximately 40 different staff EPs and 110 different residents enrolled patients.

3.2. Accuracies of MET-AP and Physicians

Table 4 outlines the overall accuracies for both triage plan prediction methods and both groups of physicians. The overall accuracy of the MET-AP triage plan recommendation (i.e., the proportion of MET-AP triage plans matching the gold standard) was very similar to the physician triage plan prediction accuracy for both groups of physicians. As anticipated, the accuracy of MET-AP recommendations based on resident assessments, as well as the resident triage plan predictions, were lower than staff EP results; however, these differences did not reach statistical significance. Further breakdown of the MET-AP triage plan recommendations and physician predictions versus the gold standard triage plans for patients seen by staff EPs is shown in Table 5, highlighting the superior performance of MET-AP for benign cases, but at the expense of poorer performance with sicker patients.

3.3. Inter-observer Agreements between Staff EP's and Residents

For the 222 patients with assessments by both a staff EP and resident, the inter-observer agreement for most attributes was moderate (0.41 to 0.60) to near perfect (0.81 to 1.00) [45], as seen in Table 6. As expected, physical exam attributes generally had lower agreement than patient history attributes, highlighting the difference in the clinical skills of the two physician groups. One attribute, localized involuntary guarding, had only fair agreement (0.21 to 0.40). Finally, moderate agreement was observed between the staff EP and resident triage plan predictions (6 = 0.575, $p < 0.001$).

4. DISCUSSION

4.1. General Remarks

The MET-AP system was developed to assist in the early management of ED patients by recommending an appropriate triage plan for all patients with AP, rather than making a definitive diagnosis or attempting to identify patients with a single diagnosis (i.e., appendicitis). With this in mind, several key principles that correlate closely with a CDSS solution's ability to improve patient care [46] were considered.

First, the solution should not interfere with the already-busy workflow in the ED. Next, it needed to focus the physician's attention, especially novices, to obtaining key historical and physical exam attributes without forcing the physician to collect all attributes. These attributes had to be limited to a reasonable number most relevant to the presentation. As such, they were selected on the advice of experienced physicians, as well as from retrospective chart data that helped to determine which attributes physicians most commonly considered. The intended use in the ED meant that triage plans were to correspond to broad groups of patients, rather than specific diagnoses, given that these management decisions must frequently be made early in the patient's ED visit before a diagnosis is established. The lack of dependence on investigations is important if MET-AP was to be used outside of a tertiary hospital ED (e.g., family physician office, community hospital) where tests may not be readily available to the healthcare provider. Finally, the system was not intended to replace the decision making of physicians by forcing the recommendation of how to manage the patient to be acted upon. Rather, we felt that MET-AP should help to reduce uncertainty by providing additional supporting evidence regarding all possible triage plans in the context of available data.

4.2. Strengths and Limitations of Our Study

To our knowledge, MET-AP is the first mobile CDSS developed for triaging pediatric patients with acute AP that can be used early in assessing a broad group of patients presenting with AP (i.e., not just a pre-selected group of patients that are more likely to have a final diagnosis of appendicitis). In addition, we are unaware of any similar CDSS providing patient-specific advice that has been tested prospectively over a prolonged period by a wide range of physician users (i.e., both expert and novice clinicians with variable levels of computer experience/comfort) in a busy teaching environment. The tests conducted with de Dombal's system, though involving several multi-centre studies, were more concerned with evaluating the decision model and structured data entry for standardized decision making, than with point of care decision support where collected data varies per presentation. The extensive participation of staff EPs and residents in our trial, as well as our success in collecting inter-observer data, allow us to confirm the validity of the attributes that were included and to see the differences between staff EPs and residents in their interpretation of these attributes.

Our study was difficult to design, as MET-AP does not provide a final diagnosis, nor estimate the probability of a single diagnosis (i.e., appendicitis) over all others. Instead, weighted recommendations of broad triage plans are provided. As such, MET-AP does not generate a single dichotomous result or continuous value on which to apply statistical methods traditionally associated with the assessment of a diagnostic test (e.g., sensitivity/specificity, receiver operator curves). For this reason, after careful consideration by our team and our statistical consultants, we chose to estimate the overall accuracy of MET-AP, reported as a percentage correct, by evaluating whether the triage plan achieving the highest weighting matched the gold standard triage plan established at follow-up. In a similar fashion, the over-all accuracy of physician predictions could be determined, and these estimates could then be compared. Further, the advisory capacity of MET-AP could not be fully appreciated due to the need to keep treating physicians blinded to the system's triage plan recommendations. Specifically, we could not evaluate how EPs would combine the weightings from all three possible triage plan recommendations with their own judgments. Once we are satisfied that MET-AP can achieve a clinically acceptable accuracy for a broad group of AP patients, we plan to evaluate patient outcomes and the impact on healthcare processes with a randomized controlled trial where these recommendations are available to physicians in the intervention arm.

We failed to reach our sample size due to difficulty maintaining physician engagement during a particularly busy viral season. Despite this, we were able to generate estimates of accuracy with sufficiently tight confidence intervals, given that the accuracies were significantly better than 50% used in the sample size calculations.

4.3. Implications and Future Research Directions

Our primary finding shows that MET-AP performs with similar overall triage plan accuracy as experienced pediatric EPs (72% vs. 70%) in determining the correct course of management early on in the patient's assessment. While this is encouraging, we are concerned with the proportion of patients that MET-AP would recommend to discharge, yet should have been further investigated or referred to the surgeon. While physicians were no more accurate overall, their triage plans were skewed towards a more conservative approach so that patients from the benign group were erroneously over-investigated or consulted.

Several factors may explain MET-AP's performance. First, MET-AP was designed for use on all patients with acute AP and not just those possibly having appendicitis, so the retrospective data contained very heterogeneous patients. In particular, patients belonging to the investigate/observe group included children with intussusception, pneumonia, urinary tract infection, and adolescents with gynecological pathology, to name a few. No single or consistent pattern of presentation is common to this diverse group of patients, resulting in weak decision rules for this group. A larger, prospective data set may yield a stronger decision model for these patients. Second, the majority of pediatric patients complaining of AP have benign conditions (e.g., gas cramps, constipation, viral illnesses), and thus should be discharged. This relative over-weighting in the prevalence of one group yields decision rules that tend to also favor this group's outcome. New data mining techniques that reduce this imbalance bias [47] are being investigated. Finally, while the attributes chosen for the system are classically quoted in the literature (i.e., they represent explicit clinical knowledge) and recorded in the charts, we may be missing more subtle observations that experienced EPs use (i.e., tacit knowledge) when deciding if a patient is sick or not. This tacit knowledge is rarely documented and may not even be recognized by the physician. We are currently conducting research using qualitative methodologies to identify and articulate this knowledge. We plan to then prospectively capture both explicit and tacit knowledge from a large, multi-centre sample of patients to be used in revising the rule-based decision model for MET-AP.

The performance of MET-AP, when used by residents, was slightly lower, though not statistically different than when used by staff EPs. This may be explained by the fact that residents may incorrectly interpret certain attributes, such as rebound tenderness or guarding, as seen in other studies [48, 49]. This results in faulty data for MET-AP to consider. The relatively lower inter-observer agreements for these attributes that require clinical skill supports this conclusion and highlights the fact that any MET-AP triage plan recommendation is only as good as the patient data that is entered. Despite not being able to compensate for inferior clinical skill, MET-AP still has an important role in prompting the less experienced physician to obtain and consider relevant historical and physical exam attributes. As seen in other studies, this alone may improve physician performance [30, 31].

The inclusion of WBC as an attribute in MET-AP is controversial, given that the decision to obtain a WBC is taking a step beyond the initial assessment that MET-AP is aimed at. Analysis of our prospective data has revealed that WBC adds limited if any discriminatory powers to the determination of the correct triage plan. As such, we intend to eliminate it from future versions of MET-AP.

We are continuing to research the rule-based decision model used in MET-AP in order to improve its overall triage plan recommendation accuracy, balanced with physicians' tendencies towards conservative management. In addition to attempting to capture and model the tacit knowledge of experts, we plan to investigate how physicians would combine the MET-AP recommendations with their own judgments in decision-making, and what impact adoption of this CDSS will have on patient outcomes. Finally, though this trial was conducted in an academic teaching department, we feel the most important settings for this type of decision support to be tested and used will be in community ED's and primary care offices, where physicians may not be as experienced and comfortable assessing children with AP. If successful in these settings, MET-AP and similar systems for other clinical problems could have more profound effects on the delivery of health services.

5. CONCLUSION

MET-AP, a rule-based clinical decision support system, recommends the correct triage plan for pediatric patients with acute AP with similar accuracy as experienced pediatrics EP's. Further research is underway to improve the accuracy and safety of this tool, leading to future evaluation of its clinical utility and acceptance by clinicians.

ACKNOWLEDGEMENTS

The authors acknowledge the assistance of Leighe-Anne Rouillard in her role as research assistant; Bernard Plouffe, Nataliya Milman and Mathieu Chiasson for their work in programming and testing the MET-AP system; Joanne Ross and the CHEO Information Services team for their help with integrating the MET-AP system with the hospital registration system; Nick Barrowman for his assistance in refining the manuscript; and the CHEO ED staff EPs and residents for recruiting patients.

This research was supported by grants from The Physicians' Services Incorporated (PSI) Foundation of Ontario and the Natural Sciences and Engineering Research Council (NSERC) of Canada (Collaborative Health Research Program).

Results of this research have been presented at the following scientific meetings: 1) Canadian Association of Emergency Physicians (CAEP) Annual Meeting, Montreal, Quebec, April 2004; 2) Pediatric Academic Societies' (PAS) Annual Meeting, San Francisco, CA, May 2004; 3) Society for Academic Emergency Medicine Annual Meeting, Orlando, FL, May 2004.

No conflicts of interest exist.

REFERENCES

References

1. T. Fathi, D. Meyer, MEDUSA: a fuzzy expert system for medical diagnosis of acute abdominal pain, Methods Inf. Med. 33 (1994) 522-529.

2. G.M. Goldman, The tacit dimension of clinical judgment, Yale J. Biol. Med. 63 (1990) 47-61.

3. J.E. Johnson, J.L. Carpenter, Medical house staff performance in physical examination, Arch. Intern. Med. 146 (1986) 937-941.

4. S. Mangione, W.P. Burdick, S.J. Peitzman, Physical diagnosis skills of physicians in training: a focused assessment, Acad. Emerg. Med. 2 (1995) 622-629.

5. K.W. Ashcroft, J.P. Murphy, R.J. Sharp, D.L. Siga, C.L. Snyder (Eds.) Pediatric Surgery, 3rd ed., WB Saunders, Philadelphia, 2000.

6. B.M. Garcia Pena, K.D. Mandl, S.J. Kraus, A.C. Fischer, G.R. Fleisher, D.P. Lund, G.A. Taylor, Ultrasonography and limited computed tomography in the diagnosis and management of appendicitis in children, JAMA 282 (1999) 1041-1046.

7. H.E. Rice, M. Arbesman, D.J. Martin, R.L. Brown, G. Gollin, J.C. Gilbert, M.G. Caty, P.L. Glick, R.G. Azizkhan, Does early ultrasonography affect management of pediatric appendicitis? A prospective analysis, J. Pediatr. Surg. 34 (1999) 754-758.

8. B.M. Pena, G.A. Taylor, D.P. Lund, K.D. Mandl, Effect of computed tomography on patient management and costs in children with suspected appendicitis, Pediatrics 104 (1999) 440-446.

9. A.B. Kharbanda, G. Taylor, S. Fishman, R. Bachur, Development of a clinical score to improve diagnosis of appendicitis and reduce use of CT scans [abstract], Pediatr. Res. 55 (2004) 598a.

10. J.W. Tsung, R. Acosta, E.F. Crain, Accuracy of the Leukocyte Count in the Diagnosis of Acute Appendicitis in Children [abstract], Acad. Emerg. Med. 9 (2002) 518a.

11. J.C. Rodriguez-Sanjuan, J.I. Martin-Parra, I. Seco, L. Garcia-Castrillo, A. Naranjo, C-reactive protein and leukocyte count in the diagnosis of acute appendicitis in children, Dis. Colon Rectum 42 (1999) 1325-1329.

12. J.M. Gronroos, Do normal leucocyte count and C-reactive protein value exclude acute appendicitis in children?, Acta Paediatr. 90 (2001) 649-651.

13. A. Alvarado, A practical score for the early diagnosis of acute appendicitis, Ann. Emerg. Med. 15 (1986) 557-564.

14. T.I. Anatol, Y. Holder, A scoring system for use in the diagnosis of acute abdominal pain in childhood, West Indian Med. J. 44 (1995) 67-69.

15. M. Samuel, Pediatric appendicitis score, J. Pediatr. Surg. 37 (2002) 877-881.

16. G.R. Bond, S.B. Tully, L.S. Chan, R.L. Bradley, Use of the MANTRELS score in childhood appendicitis: a prospective study of 187 children with abdominal pain, Ann. Emerg. Med. 19 (1990) 1014-1018.

17. C.P. Macklin, G.S. Radcliffe, J.M. Merei, M.D. Stringer, A prospective evaluation of the modified Alvarado score for acute appendicitis in children, Ann. R. Coll. Surg. Engl. 79 (1997) 203-205.

18. A.A. Malik, N.A. Wani, Continuing diagnostic challenge of acute appendicitis: evaluation through modified Alvarado score, Aust. N. Z. J. Med. 68 (1998) 504-505.

19. A.B. Kharbanda and R. Bachur, Performance of the Alvarado and Samuel appendicitis scores on a prospective cohort [abstract], Pediatr. Res. 55 (2004) 715a.

20. R.D. Goldman, S. Crump, D. Stephens, J.C. Langer, Prospective validation of a pediatric appendicitis score [abstract], Acad. Emerg. Med. 12 (2005) Suppl 1, 20.

21. M.A. Musen, Y. Shahar, E.H. Shortcliffe, Clinical decision-support systems, in: E.H. Shortcliffe, L.E. Perreault, G. Wiederhold, L.M. Fagan (Eds.) Medical Informatics. Computer Applications in Health Care and Biomedicine, 2nd edition, Springer-Verlag, 2001.

22. J.G. Anderson, Clearing the way for physicians' use of clinical information systems, Communications of the ACM 40 (1997) 83-90.

23. D.L. Hunt, R.B. Haynes, S.E. Hanna, K. Smith, Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review, JAMA 280 (1998) 1339-1346.

24. B. Kaplan, Evaluating informatics applications - Clinical decision support systems literature review, Int. J. Med. Inf. 64 (2001) 15-37.

25. M. Fieschi, J.C. Dufour, P. Staccini, J. Gouvernet, O. Bouhaddou, Medical decision support systems: old dilemmas and new paradigms?, Methods Inf. Med. 42 (2003) 190-198.

26. F.T. de Dombal, D.J. Leaper, J.R. Staniland, A.P. McCann, J.C. Horrocks, Computer-aided diagnosis of acute abdominal pain, Br. Med. J. 2 (1972) 9-13.

27. F.T. de Dombal, Surgical diagnosis assisted by a computer, Proceedings of the Royal Society of London - Biological Sciences 184 (1973) 433-440.

28. F.T. de Dombal, J.C. Horrocks, G. Walmsley, P.D. Wilson, Computer aided diagnosis and decision making in the acute abdomen, J. R. Coll. Physicians Lond. 9 (1974) 211-218.

29. G.C. Sutton, Computer-aided diagnosis: a review, Br. J. Surg. 76 (1989) 82-85.

30. I.D. Adams, M. Chan, P.C. Clifford, W.M. Cooke, V. Dallos, D. de, M.H. Edwards, D.M. Hancock, D.J. Hewett, N. McIntyre, Computer aided diagnosis of acute abdominal pain: a multicentre study, Br. Med. J. (Clin. Res. Ed). 293 (1986) 800-804.

31. P.C. Lawrence, P.C. Clifford, I.F. Taylor, Acute abdominal pain: computer aided diagnosis by non-medically qualified staff, Ann. R. Coll. Surg. Engl. 69 (1987) 233-234.

32. S. Guerlain, K. LeBeau, M. Thompson, C. Donnelly, H. McClelland, S. Syverud, J.F. Calland, The effect of a standardized data collection form on the examination and diagnosis of patients with abdominal pain [abstract], Proceedings of the Human Factors and Ergonomics Society, Human Factors an Ergonomics Society Inc., Minneapolis/St.Paul, MN, 2001, 1284-1288.

33. S. Sadeghi, A. Barzi, N. Sadeghi, B. King, A Bayesian model for triage decision support, Int. J. Med Inform. 75 (2006) 403-411.

34. W. Michalowski, S. Rubin, R. Slowinski, S. Wilk, Mobile clinical support system for pediatric emergencies, Decision Support Systems and Electronic Commerce 36 (2003) 161-176.

35. W. Michalowski, R. Slowinski, S. Wilk, K.J. Farion, J. Pike, S. Rubin, Design and development of a mobile system for supporting emergency triage, Methods Inf. Med. 44 (2005) 14-24.

36. S. Rubin, W. Michalowski, R. Slowinski, Developing an emergency room diagnostic checklist using rough sets: A case study of appendicitis. [abstract], Simulation in the Medical Sciences. Proc. of the 1996 Western Multiconference. The Society for Computer Simulation.San Diego, CA, 1996, 19-24.

37. W. Michalowski, S. Rubin, R. Slowinski, S. Wilk, Triage of the child with abdominal pain: A clinical algorithm for emergency patient management, Paediatr Child Health 6 (2001) 23-28.

38. W. Michalowski, R. Slowinski, S. Rubin, S. Wilk, Triage of acute abdominal pain in childhood: Clinical use of a palm handheld in a pediatric emergency department [abstract], Proceedings of the Hawaii International Conference on System Sciences, Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ, Big Island, HI., 2004, 2493-2499.

39. S. Wilk, R. Slowinski, W. Michalowski, S. Greco, Supporting triage of children with abdominal pain in the emergency room, European Journal of Operational Research 160 (2005) 696-709.

40. Z. Pawlak (Ed.) Rough sets: Theoretical aspects of reasoning about data, Kluwer Academic, Dordrecht, 1991.

41. S. Greco, B. Matarazzo, R. Slowinski, Dealing with missing data in rough set analysis of multi-attribute and multi-criteria decision problems, in: S.H. Zanakis, G. Doukidis, C. Zopounidis (Eds.) Decision making: Recent developments and worldwide applications, Kluwer Academic Publishers, Dordrecht, 2000.

42. S. Greco, B. Matarazzo, R. Slowinski, Fuzzy measures as a technique for rough sets analysis [abstract], Verlag Meinz, Aachen, 1998, 99-103.

43. J. Stefanowski, Classification support based on the rough sets, Foundations of Computing and Decision Sciences 18 (1993) 371-380.

44. R.G. Newcombe, Two-sided confidence intervals for the single proportion: comparison of seven methods, Stat Med 17 (1998) 857-872.

45. K.L. Posner, P.D. Sampson, R.A. Caplan, R.J. Ward, F.W. Cheney, Measuring interrater reliability among multiple raters: an example of methods for nominal data, Stat. Med 9 (1990) 1103-1115.

46. K. Kawamoto, C.A. Houlihan, D.F. Lobach, E.A. Balas, Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success, Br. Med. J. 330 (2005) 765-768.

47. J.M. Aronis, F.J. Provost, Increasing the efficiency of data mining algorithms with breadth-first marker propagation [abstract], Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, AAAI Press, 1997, 119-122.

48. J. Pines, L.U. Pines, A. Hall, J. Hunter, R. Srinivasan, C. Ghaemmaghami, The interrater variation of ED abdominal examination findings in patients with acute abdominal pain, Am. J. Emerg. Med. 23 (2005) 483-487.

49. K. Yen, A. Karpas, H.J. Pinkerton, M.H. Gorelick, Interexaminer reliability in physical examination of pediatric patients with abdominal pain, Arch. Pediatr. Adolesc. Med. 159 (2005) 373-376.

Table 1 Clinical attributes used in the MET-AP system

Attribute	Possible Values
Age	<5 years, ≥ 5 years
Gender	Male, Female
Duration of pain	≤ 24 hours, >24 hours and ≤7 days, > 7 days
Site of maximal pain	RLQ, Lower abdomen, Other
Type of pain	Continuous, Other
Shifting of pain	Yes, No
Vomiting	Yes, No
Previous visit to the ER for this AP	Yes, No
Temperature	< 37 °C, ≥37 and ≤39 °C, > 39 °C
Site of maximal tenderness	RLQ, Lower abdomen, Other
Localized involuntary guarding	Absent, Present
Rebound tenderness	Absent, Present
White blood cell count	≤ 4,000, >4,000 and <12,000, ≥ 12,000

Table 2 Categorized diagnosis list with associated triage plans

	Category: Benign/Resolving Pain		**Triage Plan: Discharge**
	Viral Syndromes		Irritable Bowel Syndrome
	Gastroenteritis		Colic, Gas Pains
	Food Poisoning		Resolved Pain Not Yet Diagnosed
	Constipation		Functional Abdominal Pain
*	Dysmenorrhoea, Mittelschmerz		
	Category: Other		**Triage Plan: Investigate/Observe**
	Hernia		Urinary Tract Infection, Pyelonephritis
	Intussusception	*	Nephrolithiasis
	Volvulus		Ovarian Torsion
	Bowel Obstruction	*	Ovarian Cyst
	Mesenteric Adenitis		Testicular Torsion
*	Gastritis, Ulcer Disease		Epididymitis
	Hepatitis	*	Diabetes
*	Biliary Colic, Cholecystitis	*	Other Metabolic Derangements
*	Pancreatitis	*	Sickle Cell Crisis
	Ingested Foreign Body	*	Henoch-Schonlein Purpura
	Intra-abdominal Neoplasm		Hemolytic Uremic Syndrome
	Leukemia	*	Porphyria
*	Inflammatory Bowel Disease		Toxin Ingestion
	Infectious Colitis		Pneumonia, Asthma
*	Malabsorptive Syndromes	*	Abdominal Wall Contusion, Strain
	Intra-abdominal Abscess (not due to ruptured appendix)		Group A Strep Pharyngitis or Tonsillitis
	Category: Appendicitis		**Triage Plan: Consult Surgery**
	Appendicitis		Intra-abdominal Abscess (due to ruptured appendix)

* Denotes conditions that may have been excluded from enrollment if a pre-existing diagnosis or recurrent process causing abdominal pain similar to the current complaint was known to the patient or physician.

Table 3 Comparison of analyzed versus not approached patient populations

Characteristic	**Analyzed** (n=574)	**Not Approached** (n=467)	p-value
Mean Age (SD)	9.13 (3.92)	9.49 (4.43)	0.170 [a]
Male n (%)	284 (49.5)	222 (47.5)	0.575 [b]
Final Triage Plan n (%)			
Discharge	443 (77.2)	335 (71.7)	0.130 [c]
Investigate/Observe	73 (12.7)	72 (15.4)	
Consult Surgery	58 (10.1)	60 (12.8)	

SD = Standard deviation
[a] Student t-test
[b] Fisher's exact test
[c] Chi-square test

Table 4 Accuracy of MET-AP triage plan recommendations and physician triage predictions by physician type

	Staff Physician Assessments (n=457)	Resident Assessments (n=339)	Difference Between Physician Type for Method
MET-AP Triage Plan Accuracy	72.2% (95% CI: 67.9, 76.1)	69.3% (95% CI: 64.2, 74.0)	2.9% p=0.755
Physician Prediction Accuracy	70.2% (95% CI: 65.9, 74.2)	62.8% (95% CI: 57.6, 67.8)	7.4% p=1.000
Difference Between Methods for Physician Type	2.0% p=0.518	6.5% p=0.836	

Table 5 MET-AP triage plan recommendation and staff EP prediction versus the gold standard triage plan established at follow-up (n=457)

		Gold Standard Triage Plan Established at Follow-up: Discharge	Investigate/ Observe	Consult Surgery
MET Triage Plan Recommendation	Discharge	279	38	12
	Investigate/ Observe	39	18	3
	Consult Surgery	30	5	33
Staff EP Prediction	Discharge	248	16	1
	Investigate/ Observe	85	39	13
	Consult Surgery	15	6	34

Shaded boxes represent patients in which the MET-AP recommendation or physician prediction correctly matches the gold standard triage plan established at follow-up.

Table 6. Inter-observer agreement between staff EPs and residents for each clinical attribute (n=222)

Attribute	**Inter-observer Agreement** (95% CI)	**Statistical Method**
Duration of pain (raw hours)	0.961 (0.949, 0.970)	ICC
Duration of pain (categorized)	0.825 (0.751, 0.900)	Kappa
Site of maximal pain	0.513 (0.409, 0.617)	Kappa
Type of pain	0.475 (0.356, 0.594)	Kappa
Shifting of pain	0.521 (0.362, 0.680)	Kappa
Vomiting	0.890 (0.829, 0.951)	Kappa
Previous visit to the ER for this AP	0.481 (0.128, 0.834)	Kappa
Temperature (raw ℃)	0.970 (0.961, 0.977)	ICC
Temperature (categorized)	0.945 (0.902, 0.989	Kappa
Site of maximal tenderness	0.573 (0.467, 0.679)	Kappa
Localized guarding	0.309 (0.139, 0.479)	Kappa
Rebound tenderness	0.449 (0.266, 0.632)	Kappa
White blood cell count (raw)	0.942 (0.891, 0.970)	ICC
White blood cell count (categorized)	0.952 (0.860, 1.000)	Kappa

Note that the attributes of age and gender were not tested for inter-observer agreement as they were automatically retrieved from the hospital registration system.
Kappa = Kappa statistic for categorical data
ICC = Intra-class correlation coefficient for continuous data

Figure 1. Example of the MET-AP triage plan recommendation screen

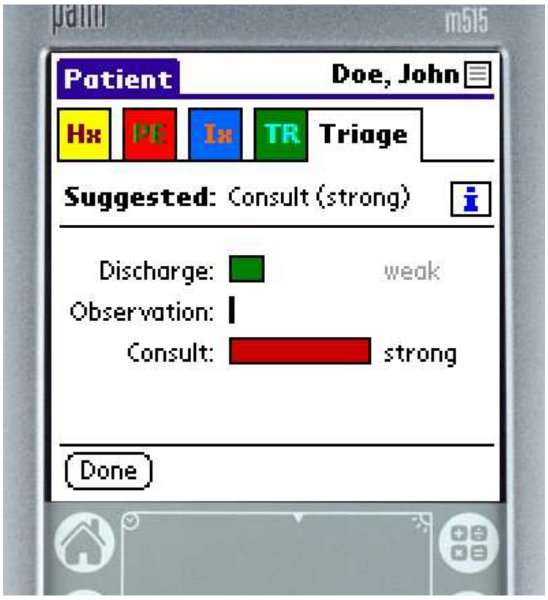

Figure 2. Data entry screen for physician's prediction of triage plan

Figure 3. Diagram describing enrollment of patients