# Decision Making By Emergency Room Physicians And Residents: Implications for the Design of Clinical Decision Support Systems.

Michael J. Hine
Eric Sprott School of Business
Carleton University
1125 Colonel By Drive
Ottawa, ON Canada, K1S 5B6
tel.: (613) 520.2600 x. 3664, fax: (613) 520-4427
e-mail: mhine@sprott.carleton.ca

Ken J. Farion
Division of Emergency Medicine
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
Telfer School of Management
University of Ottawa
136 Jean-Jacques Lussier St.
Ottawa, ON Canada, K1N 6N5
tel.: (613) 562-5800 x.4955, fax: (613) 562-5164
e-mail: wojtek@telfer.uottawa.ca

Szymon Wilk
Laboratory of Intelligent Decision Support Systems
Poznan University of Technology
Berdychowo
60-965 Poznan, Poland
tel.: +48 (61) 665-29-30, fax: +48 (61) 877-1525
e-mail: szymon.wilk@cs.put.poznan.pl

Acknowledgements: Research described in this paper was supported by grants from NSERC-CHRP and Physicians' Services Incorporated Foundation.

# Decision Making By Emergency Room Physicians And Residents: Implications for the Design of Clinical Decision Support Systems


## ABSTRACT

Clinical decision-making is complex and uncertain and is dependent on accurate and timely information that is typically managed through Information Technology (IT) solutions. One particular class of IT solution that is becoming increasingly prevalent in the medical community is Clinical Decision Support Systems (CDSS). Decision models in CDSS are typically constructed from expert knowledge and are often reliant on inputs that are difficult to obtain and are reliant on tacit knowledge that only experienced clinicians possess. Research described in this paper uses empirical results from a clinical trial of a CDSS with a decision model based on expert knowledge to show that there are differences in how clinician groups of the same specialty, but different level of expertise, elicit necessary CDSS input variables and use said variables in their clinical decisions. Using kappa analysis and logistic regression this paper reports that novice clinicians have difficulty eliciting CDSS input variables that require physical examination, yet they still use these incorrectly elicited variables in making their clinical decisions. Implications for the design of CDSS are discussed.

**INTRODUCTION**

Clinical decision-making is a complex process frequently complicated by a variety of uncertainties. It is dependent on accurate information, that according to proponents of evidence based medicine (EBM) and decision making should include the integration of clinical expertise with the best available clinical evidence generated by high quality research (Sackett, Rosenberg, Gray, Haynes & Richardson, 1996). EBM is gaining support and momentum and has been called the prevailing clinical decision making paradigm for medicine (Haynes, 2002). A need to follow EBM guidelines has resulted in a situation where clinicians are dependent on massive amounts of information and knowledge to make decisions that are in the best interest of the patient. These information and knowledge sources include electronic medical records, clinical practices guidelines, academic and practitioner journals among others. Increasingly, information technology (IT) solutions are being considered as crucial decision support mechanism to ensure that clinicians have access to appropriate knowledge sources while making clinical decisions. One particular class of IT solutions that the medical community is showing increased interest in is Clinical Decision Support Systems (CDSS).

According to a well accepted definition, a CDSS is "any program designed to help health-care professionals make clinical decisions" (Musen, Shahar & Shortcliffe, 2001). This definition includes several categories of IT solutions, including:

- *Systems for information management* that provide general data and knowledge for a variety of healthcare workers, including medical information retrieval systems for managing and extracting medical knowledge, and electronic patient record systems (EPRS: Shortcliffe, 1993) for managing patient data.

- *Systems for focusing attention* that are normally present in the intensive care units and are used to remind clinicians about actions that might require attention.
- *Systems for providing patient-specific recommendations* that assess or advise using patient-specific clinical data. These include systems ranging from direct implementation of clinical practice guidelines (Seroussi, Bouaud & Antoine, 2001) to advanced techniques of artificial intelligence (Hanson & Marshall, 2001).

CDSS from the first two categories have been relatively well accepted and used in clinical practice for more than three decades (Anderson, 1997). Increasing interest in systems from the third category is driven by a move towards EBM (deDombal, Leaper & Staniland, 1972), and the efforts to improve patient outcomes (Hunt, Haynes, Hanna & Smith, 1998). Patient-specific recommendation systems usually help clinicians make two types of decisions – diagnostic (what is the underlying health condition of the patient) and management (what is the treatment plan for the patient). Although it is rather artificial to separate the diagnostic process from the management one, many clinicians believe that it is for the management process that they would most often seek support (Musen et al., 2001).

Almost all patient-specific CDSS decision models reflect encoded clinician expertise and are reliant on accurate input to produce appropriate output that is in the best interest of the patient. The implication is that clinicians using such systems have to provide values for input variables to the CDSS that may be correctly elicited only with an appropriate level of expertise. That is, only experienced clinicians will be able to provide such information in a reliable and comprehensive manner, while inexperienced clinicians may be forced to gather information and make assessments for activities that they may lack the clinical acumen to do accurately. Thus, the resulting 'treatment plan' output provided by the CDSS may be inappropriate for the patient under question due to the poor quality of the inputs provided by the clinician.

The purpose of this paper is to challenge a common perception that a CDSS designed for a specific and well-defined clinical domain, and for users from the same domain, can satisfy the needs of clinicians who may have varying degrees of domain experience. Research described in this paper uses empirical results from a clinical trial of a CDSS to show that there are differences in how the clinician groups of the same specialty, but different level of expertise, elicit necessary CDSS input variables and use said variables in their clinical decisions. By establishing differences between the quality and use of CDSS input variables by clinicians of differing expertise we can then offer prescriptive guidance on improvements to CDSS design that ultimately should assist in providing better care to patients.

This paper is organized as follows. First, relevant background literature on expert and novice clinical decision-making is reviewed and used to formulate two research hypotheses. This is followed by a brief description of the MET-AP CDSS along with an explanation of the clinical input variables that are required by the system. Next, descriptions of the experimental design is provided, along with the analytical methodology that was used. This is followed by a discussion of the results and implications for CDSS design.

## BACKGROUND AND RESEARCH HYPOTHESES

Patient-specific CDSS are deployed in different settings and used by different classes of users. Decision models implemented in patient-specific CDSS are normally based on expert clinician knowledge, either discovered from past data, elicited from medical books or practice guidelines, or elicited directly from clinicians using a variety of knowledge acquisition strategies such as repertory grids or think aloud protocols. While techniques for obtaining expert knowledge vary, resulting patient-specific CDSS decision models almost always reflect clinician expertise. Sometimes, these models reflect ‘best practice’ by representing knowledge that has been culled from valid scientific research (for example, the encoding of a clinical practice guideline into a decision model that has been generated from systematic observations of research results). Other times, these decision models need to become part of the scientific research base from which clinicians can draw on to improve patient outcomes.

Clinicians, especially in a teaching hospital, can be considered either novice or expert, based on their medical experience and associated knowledge. Differences between these two categories of decision makers have been widely documented in the decision making and medical literature. It has been stated that in complex domains such as medicine, it typically takes 10 years of training before one can be considered an expert (Prietula & Simon, 1989). Over time, experts develop a capability to systematize information and to form complex networks of knowledge that is stored in long term memory (Arocha, Wang & Patel, 2005; Prietula & Simon, 1989). Novices lack such complex knowledge networks, and, thus, when faced with new informational cues they need to produce more hypothesis than experts (Kushniruk, 2001), are unable to filter out irrelevant cues (Patel, Arocha & Kaufman, 1994; Patel & Groen, 1991), and resultantly take a longer time in making their decisions.

In order to improve these generally weaker information gathering and decision making skills (Johnson & Carpenter, 1986; Mangione et al., 1995), medical graduates and specialty residents undergo practical training during their residency, where they learn how to assess and diagnose patients under the supervision of experienced physicians. Research has shown that residents often have deficiencies in their physical examination skills, yet they place great clinical importance on the physical examination and desire to have greater educational attention put on those skills (Mangione et al., 1995). Through self-recognizing weak skills that are widely considered critical to making important decisions, novice clinicians compensate by placing more emphasis on scientific evidence, as opposed to experts who rely on clinical experience (Patel, Groen & Patel, 1997; Patel et al., 1994). This observation was confirmed in a prospective cohort trial of a handheld CDSS for antibiotic prescribing in critical care (Sintchenko, Iredell, Gilbert & Coiera, 2005). The system offered four types of support functions: patient reports, local antibiotic guidelines, antibiotic susceptibility data and a clinical score calculator. During the trial it was observed that senior physicians used antibiotic susceptibility data more often then other support functions, while it was the least frequently used by junior physicians. The junior physician tended to use the remaining functions with local antibiotic guidelines being most frequently accessed.

Empirical studies have shown that clinicians with different levels of expertise exhibit differences in their ability to elicit information from physical examinations (Pines, Uscher Pines, Hall, Hunter, Srinivasan & Ghaemmaghami, 2005; Yen, Karpas, Pinkerton & Gorelick, 2005). In comparing abdominal examinations of Emergency Department (ED) pediatric patients undertaken by residents and attending physicians, it was found that all parts of the examination had less than moderate agreement (Yen et al., 2005). Similar results were found in studying abdominal examinations of adult patients by residents and attending physicians (Pines et al., 2005). Additional studies of residents have confirmed that they are deficient in performing physical examinations (Mangione, Burdick & Peitzman, 1995). Performing physical examinations accurately, among other clinical tasks, requires tacit knowledge that is "expressed in actions rather than conscious thoughts" (Goldman, 1990). While none of these studies involved the use of a patient-specific CDSS, the implications are that there are distinct differences between the abilities of novice and expert clinicians, and these differences may affect the novice clinicians' ability to provide accurate inputs into the expert generated CDSS decision models. The inexperienced clinicians may lack the clinical acumen necessary to make accurate elicitations and could potentially enter incorrect inputs. Such a situation may not only diminish the usefulness of the CDSS and the validity of the advice generated by the system, but also might lead to the rejection of the system by a broad group of clinicians.

The study reported here is based on a clinical trial of the Mobile Emergency Triage (MET-AP) CDSS that was developed for supporting triage decisions of pediatric abdominal pain in the ED. While the trial was originally designed to assess the CDSS's performance in terms of accuracy of the suggested decisions (Farion, Michalowski, Slowinski, Wilk & Rubin, 2004), our focus is on the CDSS decision model's input variables and the resulting decisions made by the clinicians. The decision model embedded in MET is based upon 13 input variables. We show how different clinician user groups (staff physicians (experts) and residents (novices)) used the system and made clinical decisions based on the required CDSS input variables. We also evaluate differences between these two groups and draw more general conclusions for supporting clinical decision-making with IT. Our research addresses a call for a better understanding of real decision makers making ill structured decisions in a naturalistic setting as mediated by technology (Kushniruk, 2001).

Research described here is structured around two research hypotheses. The first hypothesis builds on the results reported earlier on the differences in clinician elicitation capabilities is:

H1: Residents will not accurately elicit all values of decision making variables required by a CDSS model built from expert knowledge

It is our contention that because residents have limited clinical experience and associated tacit knowledge, they will not be able to accurately elicit values of all of the input variables for a CDSS decision model derived from expert knowledge.

The overall goal of the research described in this paper is to challenge the idea that a single CDSS is able to appropriately support clinicians of varying experience and associated expertise. To accomplish this goal we need a comprehensive assessment of both the elicitation of input variables and whether said variables are predictive of the actual decision making of clinicians of varying expertise. So while assessment of the accuracy of elicitation of CDSS input variables is critical, we are also interested in whether novice clinicians use different input variables in their clinical decisions than do staff physicians. More specifically we are concerned with whether residents rely on input variables that are relatively easy to elicit properly and that are not normally associated with clinical experience, or whether they incorporate variables that are more difficult to elicit, and traditionally require experience, into their decision making models.

Thus, our second research hypothesis is:

H2: Residents and physicians will use different decision model input variables in making their clinical decisions

Because of the clinical expertise required for certain model inputs to be correctly elicited, we expect that residents and physicians will use different input variables in their decision making models. Further, we expect that these differences will be moderated by the 'type of input variable', with variables requiring tacit knowledge and clinical experience to be less important in residents decision making models. This would be consistent with classical decision making where it is stated that decision makers will use the best information available and if there is uncertainty, the decision makers will act in a way to reduce uncertainty if possible (Simon, 1957).

**CDSS: MET-AP**

The MET-AP CDSS was designed and developed to support ED clinicians in making triage decisions about children with abdominal pain (Michalowski, Slowinski, Wilk, Farion, Pike & Rubin, 2005). It facilitates early patient management by ED clinicians who need to make decisions about the clinical management of patients based on initial clinical history and assessment. In this sense MET-AP is not a diagnostic CDSS because it does not provide clinicians with a differential diagnosis but rather with broad management categories (i.e. discharge from the ED, keep for further observation, or request specialist consult). The MET-AP system architecture consists of a server that interfaces with the hospital's electronic patient record system using the HL7 protocol (Quinn, 1999) and clients that reside on mobile devices such as a Personal Digital Assistant (PDA). The client facilitates the collection of clinical data (CDSS input variables) at the point of care and is used during patient examination by the physician.

The system provides a user interface composed of a series of screens to collect 11 out of 13 CDSS input variables required by the pediatric abdominal pain triaging model. These include variables related to physical findings as well as patient history. The remaining two variables, gender and age, are extracted automatically from the electronic patient record system. All variables are detailed in Table 1 and were identified using retrospective chart analysis. The triage decision making model was created using knowledge discovery techniques based on rough set theory (Pawlak, 1991; Slowinski, 1995) and implemented as a rule-based model.

Based on the values of the input variables the MET-AP's triage model generates suggested triage decision which can be one of the following three outcomes:

- *Discharge*: patient can be discharged home as his/her pain is caused by a non-serious problem,

- *Observation/Investigation*: further in-hospital evaluation (either in the ED or hospital ward) is required to better evaluate the cause of the pain,
- *Consult*: surgical consult is required due to suspicion of acute appendicitis (most common surgical emergency in children with abdominal pain).

**Table 1:** Abdominal pain triaging attributes

Insert Table 1 Here

Values of all numerical input variables (WBC, temperature, duration of pain) were collected by physicians entering direct numerical values using either a virtual keyboard or a handwriting recognition system. Entered values were then discretized for the rule based decision models according to discretization norms developed with physician experts. Values of all input variables that involved a specific location within the abdomen (site of maximal pain, site of maximal tenderness) were collected by physicians tapping on an abdomen pictogram on the mobile device. Other input variables were collected via standard user interfaces for mobile devices. For example, figure 1 shows the MET-AP screen for 'type of maximal pain'. All screens were designed and developed with participation from multiple physicians. This ensured that the resulting user interface mimicked clinicians' natural data collection procedures as closely as possible.

**Figure 1:** MET-AP screen for type of pain

Insert Figure 1 Here

## METHODS

This research on staff physician and resident decision making was part of a larger clinical trial that was designed to evaluate MET-AP decision accuracy in comparison with clinicians' triage predictions. Results of that clinical trial can be found in Farion, Michalowski, Rubin, Wilk, Correll and Gaboury (2008).

### Sample and Data Collection

A convenience sample of 574 eligible children with acute abdominal pain, aged 1 to 16 years, were enrolled with consent between July 2, 2003 and February 29, 2004 in the ED of the Children's Hospital of Eastern Ontario, Ontario, Canada. Under some conditions, convenience samples are not representative of the population under study (in this case, children with acute abdominal pain). While there were a variety of factors that effected enrolment of patients, including how busy the ED was and the attending clinician's level of comfort with technology, because of the long enrolment period (8 months), the number of clinicians involved (150), and the number of patients seen (574), it is likely that the patient sample is reasonably representative of the population under study.

A typical MET-AP usage scenario for clinicians participating in the study is presented in figure 2. After logging in to the MET-AP system, the attending resident or staff physician would enroll a patient and collect and record their medical history into the system. The attending clinician would then typically collect physical findings from the patient through physical examination and verbal interaction and enter the relevant input variables into the MET-AP system via the user interface. Participating clinicians were instructed to only record data for those input variables they felt were relevant to the patient's presentation. After reflecting on the findings the attending clinician, blinded to the CDSS recommendation, entered his/her prediction of which triage category the patient was most likely to fit (i.e., discharge, observation/investigation, or consult). Where possible, a clinician with a different level of expertise (i.e., resident or staff physician) from the attending clinician was asked to complete an independent interrater assessment within one hour of the original assessment using the MET-AP system as described above.

**Figure 2:** MET-AP usage scenario

Insert Figure 2 Here

Forty staff physicians and one hundred and ten residents enrolled patients. This type of prospective evaluation of CDSS is rare, as all physicians were asked to use the MET-AP, not just those few associated with the development team. The physicians had varying degrees of experience with handheld computers before entering the trial and all of them participated in in-depth training sessions after which they were able to easily use the CDSS. Two hundred and twenty two of the patients were seen by both a resident and a staff physician.

**Analysis**

The analysis starts by addressing H1 to establish whether residents are accurate at eliciting the input variables required by MET-AP. Once that is established, H2 is addressed to compare which MET-AP input variables predict the triage decision in the resident and staff physician decision making models. In this study, the more experienced staff physicians' inputs represent the benchmark to which residents' values are compared. This approach is widely used in the literature to evaluate performance of less experienced clinicians and can take the form of comparing novices to experts performing the same task (Nodine, Kundel, Mello-Thoms, Weinstein, Orel, Sullivan & Conant, 1999; Sklar, Hauswald & Johnson, 1991), or having expert clinicians evaluating the performance of novice clinicians (Burdick et al., 1996; Steinbach, 2002; Wray & Friedland, 1983). As a measure of proper elicitation, we use a level of agreement beyond chance between values for CDSS input variables provided by staff physicians and residents. Statistically, this is measured using the Cohen's Kappa statistic (Cohen, 1960) which was calculated for each of input variables across the two groups of clinicians who have seen the same patient.

Addressing H2 involves the use of logistic regression to determine which CDSS input variables are significant in predicting the clinicians' triage decision. In this analysis, the CDSS input variables are independent variables, and the triage decision made by staff physicians and residents is the dependent variable. It should be clear that the dependent variable is the clinician's actual triage decision, not the decision provided by the CDSS. Logistic regression was chosen given that the dependent variable was categorical. In conducting this analysis we collapsed the original three possible values for the dependent variable (the clinician's triage decision) into two distinct values. This was done by combining 'observation/investigation' and 'discharge' into one category. 'Consult' remained a distinct category. This isolated the significance of the input variables associated with the 'consult' value of the dependent variable. This situation serves as a proxy for a critical triage decision typical of a diagnosis of acute appendicitis.

The regression analysis was conducted separately for data derived from patients who were seen by residents, and patients who were seen by staff physicians so we could investigate and compare decision making models across clinician type. Typical model building strategies suggest doing extensive univariant analyss for each potential independent variable to determine which variables should be added to the model (Hosmer & Lemeshow, 2000). However, epidemiological researchers suggest including all clinically and intuitively relevant variables into the initial model regardless of their significance. Because the input variables included in the MET-AP were derived from a retrospective chart study and were validated with ED physicians, all of them were included in the analysis. Before running the regression, we studied the contingency tables for all independent variables against the dependent variable to ensure that no zero cells existed. This basic logistic regression requirement was met successfully for both resident and staff physician data.

*Design Effect*

Because this study involved a prospective trial in the ED, it was unrealistic to obtain random sampling of patients, residents and staff physicians. In situations like this, the cluster sampling of a population may suffer from a sampling bias. In order to determine if this is the case, design effects (DEFF) are calculated. This measure assumes that the respondents in the same cluster are likely to be similar to one another and thus each respondent from a cluster typically contributes less new information than would a randomized respondent. The DEFF is calculated as a ratio of the variance under the sampling method employed to the variance computed using simple random sampling (Skinner, Holt & Smith, 1989):

DEFF = 1 + $\delta$ (n-1), where:

$\delta$ is the intercluster correlation for the statistic in question

n is the average size of the cluster

The sample used in our study is not independent because there were multiple staff physicians and multiple residents, each of whom saw more than one patient. A cluster was formed by grouping together the multiple patients seen by a given staff physician and the multiple patients seen by a single resident. Because information on the performance of individual clinicians was not permitted by the Research Ethics Board, the association between staff physician/resident to individual patients is unavailable, and thus it is impossible to calculate $\delta$ and subsequently DEFF. To alleviate the concern around clustering, we calculated a ‘critical DEFF’ defined as the DEFF that would adjust the statistic in question to the point where it was no longer significant at value of 0.05. This approach has been used successfully in previous research (Thomas & Cyr, 2002). The critical DEFF was calculated as:

Critical DEFF = $\frac{\hat{W}}{c^2}$

Where:

$\hat{W}$ is the calculated Wald Statistic for the CDSS input variable in question, and $c^2$ is the critical chi square value for n-1 degrees of freedom. While values of DEFF can vary depending on the study design and individual variable in question, research suggests that a well-designed study should result in values of DEFF between 1 and 3 (Shackman, 2001). While it is impossible to accurately estimate the DEFF for this study, we would expect its value to be very low. For the physicians clustering, we would not expect the likelihood that a randomly selected staff physician from the overall population would provide input variable values much different from those currently elicited. While we expect there would be higher variance for the residents (because of less expertise), the cluster size for the resident population in the study is small (because of the large number of residents participating) which might contribute to a lower value of DEFF.

## RESULTS

### H1: Accuracy of Collected Inputs

Kappa measures and associated interpretation information (Posner, Sampson, Caplan, Ward & Cheney, 1990) of agreement between staff physicians and residents for CDSS input variables are presented in table 2. It should be noted that all of the input variables were assessed using discretized values. While some of the input variables are naturally scalar data (for example, temperature), the discretizations adopted were generated by a panel of experts and reflect critical threshold as used by clinicians in daily practice. As expected, input variables which are objective and easily measured or assessed (vomiting, temperature) have high levels of agreement indicating that residents are able to accurately elicit this information. However, all other input variables had only fair or moderate levels of agreement. Except for ‘previous visit’, the elicitation of these input variables are more difficult and subjective than the previous mentioned input variables. ‘Previous visit’ is defined as a “previous visit to the ED for abdominal pain during the last 48 hours (irrespective of site)”. We suspect that the low value of the Kappa statistic may be attributed to the fact that some of the patient/parent(s) interpreted the first examination (conducted by staff physician/resident) as a previous visit when they were asked the same question by the second observer.

**Table 2:** Kappa statistic: Resident vs. Physician

| |
|---|
| Insert Table 2 Here |

The values of Kappa statistic indicate that residents are less accurate eliciting input variable values that require experience and clinical acumen, as opposed to straight application of 'textbook knowledge'. Of the input variables that had fair to moderate levels of agreement, the ones with the lowest values of Kappa (localized guarding and rebound tenderness) are more dependent on experience in conducting physical examination than the remaining attributes (type of pain, site of pain, shifting of pain, site of tenderness) and are typically considered the most difficult to accurately elicit. The elicitation of these physical examination input variables can be obstructed due to the child's sensitivity to being touched, his/her fear, and other factors that may cause muscle contraction leading to misinterpretation. The examination for rebound tenderness is painful for patients when it is present, so repeated examinations to confirm this finding is discouraged. Thus, experience in carrying out examinations is likely to increase the reliability of eliciting values for physical examination input variables. Residents may not have enough experience to distinguish the subtle difference between a patient with true guarding and one that is just uncomfortable with the physical examination (Mangione et al., 1995). At the same time it is important to recognize that according to clinical knowledge, the combination of the presence of localized guarding and rebound tenderness is a 'strong indicator' for surgical consult due to possible appendicitis. In the case of MET-AP input variables, those that are the most difficult for residents to elicit provide the most insight into the patient's state. In summary, those attributes that required physical examination, and thus clinical acumen and experience, to accurately elicit their values for CDSS inputs were done poorly by residents.

The remaining input variables having moderate level of agreement (type of pain, site of pain, shifting of pain, and site of tenderness) are reliant on the ability of the physician to 'touch and ask' to elicit accurate values from the patient. The capability to elicit an accurate response through the dynamic interplay between clinician and patient is affected by level of expertise, with less experienced physicians having weaker information gathering skills (Johnson & Carpenter, 1986; Mangione et al., 1995).

Based on these results H1 is supported. The results add further evidence to the literature that residents do not have sufficient clinical expertise required to reliably elicit information that is dependent on the physical examination. In the next step of our research we wanted to determine differences between MET-AP input variables used by residents and staff physicians in making their triage decisions. Because of the clinical experience required for certain input variables to be correctly elicited, we expect that residents and physicians will use different input variables in their mental decision making models.

**H2: Critical Decision Making Variables**

The results for residents and staff physicians are shown in Tables 3 and 4 respectively. The values of Nagelkerke's $R^2$ is 0.568 and 0.699 for the resident and staff physician models indicating that the CDSS input variables provide a better fit for staff physician mental model than for the resident model.

The design effects are reflected in critical DEFF values that are shown in Tables 3 and 4 to the immediate right of the calculated p-values. For example, for the ‘localized guarding’ input variable for the resident analysis, a critical DEFF value of 3.508 is the minimal value required to categorize ‘localized guarding’ as insignificant. Based on the critical DEFF values for the physician analysis, we would expect one of the ‘significant variables’ to become insignificant if simple random sampling was used. Specifically, vomiting (with a critical DEFF of 1.217) will most likely become insignificant. The critical DEFF values for the input variables that are significant for the residents’ model are all high enough to expect that these input variables would remain significant if randomized sampling was used. Because of the difficulties associated with calculating values of DEFF and a need to resort to using critical DEFF instead, the results presented here should be interpreted with caution. While this could be viewed as a limitation it should be noted that prospective trial data of CDSS use in a naturalistic setting is rare and efforts should be taken to use the data in a responsible academic fashion.

**Table 3:** Logistic regression for residents

Insert Table 3 Here

**Table 4:** Logistic regression for staff physicians

Insert Table 4 Here

Localized guarding and rebound tenderness are highly significant in the residents’ mental model for making the consult decision. This is not surprising considering the importance of these input variables in determining acute appendicitis. When taken in concert with the Kappa statistic for the same input variables we have a situation where residents rely on input variables that they have trouble eliciting correctly while making a decision to ask for a surgical consult. There are several plausible diverging explanations for this result dependent upon whether residents are cognizant of their (in)ability to properly elicit certain input variables.

Residents may be aware of their deficiencies in eliciting certain input variables, but that awareness is counter balanced or overridden by an accepted clinical guideline that is reliant on aforementioned variables. While this seems counter-intuitive, it is not entirely inconsistent with previous research where residents have been shown to be deficient at performing physical examinations, yet they acknowledge both their own deficiencies and the importance of being able to properly and reliably do physical examinations (Mangione et al., 1995). Our results indicate that the perceived importance of localized guarding and rebound tenderness by the residents outweighs their perception of the degree of difficulty in eliciting the input variable values and the degree of reliability in collecting this information, and thus they use these input variables when making a consult triage decision.

Alternatively, residents may incorrectly feel confident in their ability to elicit all input variables required by the CDSS decision model. In this case we would expect the residents to consider and apply the variables as specified in their training and education into decision making activities. Previous studies have shown that residents cannot accurately estimate their performance and that they have a tendency to overestimate their performance (Parker, Alford & Passmore, 2004). It has also been shown that while residents and physicians both overestimate the accuracy of their clinical diagnoses, residents overestimate more often than physicians (Friedman, Gatti, Franz, Miller & Elstein, 2005). More research is required to investigate perceptions of accuracy of CDSS input variables and resulting clinical decisions.

Overall, staff physicians have more significant variables that predict their triage decision making than do residents. Specifically, site of pain, type of pain, localized guarding and rebound tenderness are significant predictors for physicians. Alternatively, significant predictive variables in the residents' model are site of tenderness, localized guarding and rebound tenderness. H2 is thus supported. These results are consistent with the literature on strategic experts, which states that experts have complex structures that assist in the recognition and interpretation of environmental signals and events (Lyles & Schwenk, 1992) and that these structures are more complex and contain more links among elements than the cognitive structures of less experienced strategists (Day & Lord, 1992; Lurigio & Carrol, 1985; McKeithen, Reitman, Rueter & Hirtle, 1981).

## LIMITATIONS, CONCLUSIONS AND IMPLICATIONS FOR CDSS DESIGN

There are several limitations associated with this study that are worth mentioning. First, a convenience sample of patients was used which can limit the generalizability of the results as there is no guarantee that enrolled patients were representative of the overall population of interest (children with abdominal pain). An additional limitation was the time lag between the assessments made by the clinicians of differing expertise on the same patient (< 1 hour). It is possible in some cases that the patient's condition could change during the time between assessments. There was nothing to indicate that this effect was occurring during the data collection process. As mentioned previously, the sample used in our study is not independent because there were multiple staff physicians and multiple residents, each of whom saw more than one patient. Without the knowledge of which physician and which resident saw which patients, we were unable to apply more advanced statistical techniques such as hierarchical linear modeling to determine the decision making models of the different clinician groups. However, we did attempt to alleviate the above problem by calculating critical DEFF values and applying said values to refine the final logistic regression models. A final limitation relates to the enrollment of patients by the attending clinicians. Those clinicians who were less comfortable with the MET-AP technology could be less likely to enroll patients into the trial.

The quality of any patient specific CDSS is reliant on the quality of the underlying decision model(s). These models have to reflect clinical expertise associated with expert decision makers (staff physicians in our situation). Models associated with such expertise will usually require inputs that are difficult to assess and interpret by novice users. Broadly speaking, customizing CDSS technology for users of different expertise has been proposed by several researchers (Kushniruk, 2001; Patel, Arocha, Diermeier, How & Mottur-Pilson, 2001), but to our knowledge this research is one of the first that provides empirical evidence gathered through the prospective evaluation of a CDSS, that such an approach is required. In typical CDSS designs, residents and physicians would be treated as a single user group, and thus would be interacting and accessing the same interface and underlying decision models.

In evaluating the use of a CDSS for ED triage of patients with abdominal pain, we found that staff physicians and residents elicited several of the CDSS input variables differently while examining the same patients. Specifically, for CDSS input variables involving physical examination typically in concert with verbal elicitation, calculated Kappa values were low indicating that the values recorded by the residents were different than those recorded by staff physicians. Considering that we use staff physicians' values as the benchmark (in accordance with expert vs novice literature), we interpret this discrepancy as indicative of the difficulties the residents had with correctly eliciting values of such input variables. When individual mental decision making models (operationalized as significant CDSS input variables for predicting triage decisions as a proxy) were examined it was found that the staff physicians and residents models were similar, albeit the staff physicians' model had one more significant variable. The importance of the input variables that required physical examination was underlined by their presence in both staff physicians' and residents' mental models, even though the residents were not eliciting this information accurately. In order to take into account differences in clinical experience and to ensure appropriate support is available to these different user groups, we propose that the CDSS designers should (a) differentiate between information values provided by the data coming from expert and novice assessments, and (b) implement logical attribute monitoring that warns users when a single attribute value or a combination of attribute values is outside of expected ranges or patterns.

To design and implement aids that consider the information value of the inputs, the input variables used in CDSS models must be categorized. Required input variables could be logically categorized based on how difficult they are to elicit and to what extent they are reliant on tacit, explicit, or declarative knowledge. Subsequently, each input variable could be indicated as 'low confidence' or 'high confidence'. While this is a broad categorization, it reflects the ability of different physician user groups to accurately elicit different values of the input variables. While the categorization of the variables is encoded into the system, it can remain relatively transparent to the user (i.e., there would be nothing that would explicitly label a variable as being 'low confidence'). According to the proposed categorization, a typical novice physician would have elicitation difficulty with 'low confidence' input variables. Therefore, the user interface for the 'low confidence' attributes should provide extensive explanations and guidelines to assist the process of collection. Some progress has been made in providing explanations and guidelines for CDSS input elicitation. AI/RHEUM (Kingsland, Lindberg & Sharp, 1983) is an expert-based system for diagnosing rheumatic diseases and was created to provide knowledge elicited from rheumatology experts to physicians with no training in rheumatology. To support physicians in providing accurate input information, AI/RHEUM included an extensive repository of 180 definitions of items from the finding list (Porter, Kingsland, Lindberg, Shah, Benge, Hazelwood, Kay, Homma, Akizuki, Takano & Sharp, 1988). A more recent version of the system this information was augmented with multimedia presentations including videos and pictures and a function to search for referenced articles directly on Medline (Athreya, Cheh & Kingsland, 1998).

Provision for recording imprecise or uncertain information (e.g., selecting several values instead of a single one, entering some 'confidence' factor associated with a value, or having a discrete option for 'uncertain') should be provided. Additional factors related to the process of eliciting values for 'low confidence' input variables should be considered by expanding the clinical value set with conditional information such as ‘recorded with difficulty’, or ‘child crying and fidgeting’. This will allow a dynamic confidence factor to be calculated. Moreover, to help with ‘learning by analogy’, at any time, and at the users discretion, similar patient cases could be retrieved based on values of individual input variables or on a more complete clinical model. This approach is consistent with knowledge transfer literature that states that while tacit knowledge cannot necessarily be made explicit, it can be transferred through repeated exposure to similar situations and cases (Nickols, 2000). ‘High confidence’ input variables would not require such additional assistance and could be elicited in the usual manner. Finally, following accepted principles of interaction design, the additional input support functionality for low confidence attributes discussed above should be automatically turned on for less experienced clinicians, while more experienced clinicians could bypass the additional support if desired (Shneiderman, 1998).

In clinical decision making, values of selected attributes often form a certain pattern that is indicative of an underlying health condition. For example, as stated earlier, for pediatric abdominal pain, pain and tenderness located in the right lower quadrant in concert with presence of guarding are indicative of possible acute appendicitis. It is possible to use information about such patterns to develop context sensitive monitoring for values of both individual input variables and their combinations. If values entered by a clinician significantly deviate from the dynamically adjusting thresholds, either assessed individually or within clinical patterns, a CDSS would issue specific warning alerting the clinician to this situation. While this will provide additional support for novice clinicians, it will also help minimize the potential error between user and technology which has recently been identified as an important source of clinical error

(Kohn, Corrigan & Donaldson, 2000). The derivation of the thresholds of the input variables should be generated dynamically based on an abstraction of the patient profile and subsequent heuristic matching against a set of likely profiles developed on a basis of past cases. The case base could provide the core knowledge repository on which to derive the threshold values that can be obtained in a manner similar to case-based reasoning in artificial intelligence. Machine learning algorithms and induction techniques could also be adopted to derive threshold values, rules, and patterns that new patient profile information can be compared to. These approaches assume a sufficiently large case database to ensure realistic variances are reflected in establishing the threshold values.

Many decision models implemented into CDSS encapsulate knowledge that relies on evaluating attributes that require experience and significant clinical acumen. Results of the research reported here indicate that residents have not completely mastered this knowledge and thus encounter difficulties with providing the required input to the CDSS. This creates uncertainty about the quality of the recommendations produced by the CDSS. It is clear that customized decision support, taking into account the level of clinical expertise and background of a given physician, is required to ensure that the accuracy of the CDSS is maximized. Such expanded support is as important for the acceptance of a CDSS by physicians as the quality of the underlying decision model and user interface.

REFERENCES

Anderson, J.G. (1997). Clearing the way for physicians' use of clinical information systems. *Communications of the ACM, 40*(8), 83-90.

Arocha, J.F., Wang, D. & Patel, V. (2005). Identifying reasoning strategies in medical decision making: A methodological guide. *Journal of Biomedical Informatics, 38*(2), 154-171.

Athreya B.H., Cheh, M.L., Kingsland, L.C. (1998). Computer-assisted Diagnosis of Pediatric Rheumatic Diseases. *Pediatrics,* 102 (4), e48.

Burdick, W.P., Ben-David, M.F., Swisher, L., Becher, J., Magee, D., McNamara, R., Zwanger, M. (1996). Reliability of performance-based clinical skill assessment of emergency medicine residents. *Academic Emergency Medicine,* 3(12), 1119-23.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46.

Day, D.V. & Lord, A.G. (1992). Expertise and problem categorization: The role of expert processing in organizational sense-making. *Journal of Management Studies,* 29(1), 35-47.

de Dombal, F.T., Leaper, D.J., Staniland, J.R., McCann, A.P. & Horrocks, J.C. (1972). Computer-aided diagnosis of acute abdominal pain. *British Medical Journal*, *2*(5804), 9-13.

Farion, K.J., Michalowski, W., Rubin, S., Wilk, S., Correll, R. & Gaboury, I. (2008). Prospective evaluation of the MET-AP system providing triage plans for acute pediatric abdominal pain. *International Journal of Medical Informatics,* 77(3), 208-218.

Farion, K.J., Michalowski, W., Slowinski, R., Wilk, S. & Rubin, S. (2004). Rough set methodology in clinical practice: Controlled hospital trial of the MET system. In S. Tsumoto, R. Slowinski, J. Komorowski & J. Grzymala-Busse (Eds.), *Rough sets and current trends in computing*. Heidelberg: Springer Verlag, 805-814.

Friedman, C.P., Gatti ,G.G., Franz, T.M., Murphy, G.C., Wolf, F.M., Heckerling, P.S., Fine, P.L., Miller, T.M., Elstein, A.S. (2005). Do physicians know when their diagnoses are correct? Implications for decision support and error reduction. *Journal of General Internal Medicine, 20*(4), 334-339.

Goldman, G.M. (1990). The tacit dimension of clinical judgment. *The Yale Journal of Biology and Medicine*, 63, 47-61.

Hanson 3rd, C.W. & Marshall, B. E. (2001). Artificial intelligence applications in the intensive care unit. *Critical Care Medicine*, *29*(2), 427-35.

Haynes R.B. (2002). What kind of evidence is it that Evidence-Based Medicine advocates want health care providers and consumers to pay attention to? *BMC Health Services Research, 6*(2), 1-7.

Hosmer, D.W., & Lemeshow, S. (2000). *Applied logistic regression (2nd ed.).* New York, NY: Wiley Interscience.

Hunt, D.L., Haynes, R.B., Hanna, S.E. & Smith, K. (1998). Effects of computer-based clinical decision support systems on physician performance and patient outcomes: A systematic review. *The Journal of the American Medical Association*, *280*(15), 1339-1346.

Johnson, J.E. & Carpenter, J.L. (1986). Medical house staff performance in physical examination. *Archives of Internal Medicine, 146(5)*, 937-941.

Kingsland, L.C., Lindberg, D.A.B., Sharp, G.C. (1983). AI/RHEUM. A Consultant System for Rheumatology. *Journal of Medical Systems*, 7 (3), 221–227.

Kohn, L.T., Corrigan, J.M. & Donaldson, M.S. (2000). *To err is human: building a safer health system*. Institute of Medicine. National Academy Press: Washington, DC.

Kushniruk, A.W. (2001). Analysis of complex decision-making processes in health care: Cognitive approaches to health informatics. *Journal of Biomedical Informatics*, *34*(5), 365-376.

Lurigio, A.J. & Carrol, J.S. (1985). Probation officers' schemata of offenders: Content, development and impact of treatment decisions. *Journal of Personality and Social Psychology, 48*(5), 1112-1126.

Lyles, M.A. & Schwenk, C.R. (1992). Top management, strategy and organizational knowledge structures. *Journal of Management Studies*, *29*(2), 155-174.

Mangione, M., Burdick, W.P. & Peitzman, S.J. (1995). Physical diagnosis skills of physicians in training: a focused assessment. *Academic Emergency Medicine, 2*, 622-629.

McKeithen, K.B, Reitman, J.S., Rueter, H.H., & Hirtle, S.C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, *13*(3), 307-325.

Michalowski, W. R. Slowinski, Sz. Wilk, K. Farion, J. Pike and S. Rubin (2005) Design and Development of a Mobile System for Supporting Emergency Triage. *Methods of Information in Medicine 44*(1) 14-24.

Musen, M.A., Shahar, Y. & Shortliffe, E.H. (2001). Clinical decision support systems. In E. H. Shortliffe, L. E. Perreault, G. Wiederhold, & L. M. Fagan (Eds.), *Medical informatics. Computer applications in health care and biomedicine (2 ed.).* New York, NY: Springer, 573-609.

Nickols, F.W. (2000). The knowledge in knowledge management. In J.W. Cortada & J.A. Woods (Eds.), *The knowledge management yearbook 2000-2001*. Boston, MA: Butterworth-Heinemann, 12-21.

Nodine C.F., Kundel H.L., Mello-Thoms C., Weinstein S.P., Orel S.G., Sullivan, D.C. & Conant E.F. (1999). How experience and training influence mammography expertise. *Academic Radiology*, *6*(10),575-85.

Parker, R.W., Alford C. & Passmore C. (2004). Can family medicine residents predict their performance on the in-training examination? *Family Medicine, 36*(10), 705-709.

Patel, V., Arocha, J.F., Diermeier, M., How, V. & Mottur-Pilson, C. (2001). Cognitive psychological studies of representation and use of clinical practice guidelines. *International Journal of Medical Informatics, 63*(3), 147-167.

Patel, V.L., Arocha, J.F. & Kaufman, D.R. (1994). Diagnostic reasoning and medical expertise. *Psychology of Learning Motivaion*, *31*, 187-252.

Patel, V.L., Groen, G.J. & Patel, Y.C. (1997). Cognitive aspects of clinical performance during patient workup: The role of medical expertise. *Advances in health sciences Education*, *2*(2), 95-114.

Patel, V.L. & Groen, G.J. (1991). The general and specific nature of medical expertise. *In: Ericsson, A. & Smith, J. editors. Toward a general theory of expertise: prospects and limits.* Cambridge Univ. Press, 93-125.

Pawlak, Z. (1991). *Rough sets: Theoretical aspects of reasoning about data*. Norwell, MA: Kluwer Academic Publishers.

Pines, J., Uscher Pines, L. Hall, A., Hunter, J., Srinivasan, R., & Ghaemmaghami, C. (2005). The interrater variation of ED abdominal examination findings in patients with acute abdominal pain. *American Journal of Emergency Medicine*. *23*, 483-487.

Porter, J.F., Kingsland, L.C., Lindberg D.A.B., Shah, I., Benge, J.M., Hazelwood, S.E., Kay, D.R., Homma, M., Akizuki, M., Takano, M., Sharp, G.C. (1988). The AI/RHEUM Knowledge-based Computer Consultant System in Rheumatology. Performance in the Diagnosis of 59 Connective Tissue Disease Patients from Japan. *Arthritis and Rheumatism,* 31 (2), 219–226.

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

Prietula, M.J. & Simon, H. (1989). The experts in your midst. *Harvard Business Review, 67*(1), 120-124.

Quinn, J. (1999). An HL7 (health level seven) overview. *Journal of American Health Information Management Association, 70*(7), 32-34.

Sackett, D.L., Rosenberg, W.M., Gray, J.A., Haynes, R.B. & Richardson, W.S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal, 312*, 71-72.

Séroussi, B., Bouaud, J., & Antoine, É.C. (2001). OncoDoc: A successful experiment of computer-supported guideline development and implementation in the treatment of breast cancer. *Artificial Intelligence in Medicine, 22*(1), 43-64.

Shackman, G. (2001). Sample size and design effect. *presented at Albany Chapter of American Statistical Association*.

Shneiderman, B. (1998). *Designing the user interface (3rd ed.)*, Addison Wesley.

Shortliffe, E.H. (1993). The adolescence of AI in medicine: Will the field come of age in the 90's? *Artificial Intelligence in Medicine, 5*(2), 93-106.

Simon, H. (1957). *Administrative behavior (2nd ed.)*, New York: Free Press.

Sintchenko, V., Iredell, J.R., Gilbert, G.L., Coiera, E. (2005). Handheld Computer-based Decision Support Reduces Patient Length of Stay and Antibiotic Prescribing in Critical Care. *Journal of the American Medical Informatics Association*, 12 (4), 398–402.

Skinner, C.J., Holt, D., & Smith, T.M.F. (1989). *Analysis of complex surveys*. Chichester, NY: Wiley.

Sklar, D.P., Hauswald, M., Johnson, D.R. (1991). Medical problem solving and uncertainty in the emergency department. *Annals of Emergency Medicine, 20(9)*, 987-91.

Slowinski, R. (1995). Rough sets approach to decision analysis. *AI Expert Magazine, 10*, 8-25.

Steinbach, W.J., Sectish, T.C., Benjamin, D.K. Jr, Chang, K.W., Messner, A.H. (2002). Pediatric residents' clinical diagnostic accuracy of otitis media. *Pediatrics*, 109(6), 993-998.

Thomas, D.R. & Cyr, A. (2002). Applying item response theory methods to complex survey data. *Proceedings of the Survey Methods Section, Statistical Society of Canada Annual Meeting*, Hamilton, ON, Canada

Wray, N.P. & Friedland, J.A. (1983). Detection and correction of house staff error in physical diagnosis. *Journal of the American Medical Association*, 249(8), 1035-1037.

Yen, K., Karpas, A., Pinkerton, J.J., & Gorelick, M.H. (2005). Interexaminer reliability in physical examination of pediatric patients with abdominal pain. *Archives of Pediatrics & Adolescent Medicine, 159*, 373-376.

**Table 1:** Abdominal Pain Triaging Attributes

| Attribute Name & Description | Possible Values |
|---|---|
| Age | 0-5, >5 years |
| Localized guarding: localized muscle sustained contraction noted when palpating the abdomen | Absent, Present |
| Duration of pain | <=24 hrs, 1-7 and >7 days |
| Shifting of pain | Absent, Present |
| Site of maximal pain | Right lower quadrant (RLQ), lower abdomen, other |
| Type of maximal pain | continuous, other |
| Previous visits in the ED for abdominal pain during the last 48 hours (irrespective of site) | yes, no |
| Rebound tenderness: pain felt at site of maximal tenderness, produced by altering intra-abdominal pressure | absent, present |
| Gender | male, female |
| Temperature | <37, 37-39, > 39 Cel |
| Site of maximal tenderness | RLQ, lower abdomen, other |
| Vomiting | yes, no |
| WBC (white blood cells) | <=4000, 4000-12000, >=12000 |

**Table 2:** Values of Kappa Statistic: Resident vs. Staff Physician

| Attribute | Kappa | Agreement Quality[1] |
|---|---|---|
| Localized guarding | 0.31 | Fair |
| Rebound tenderness | 0.45 | Moderate |
| Previous visit | 0.48 | Moderate |
| Type of pain | 0.48 | Moderate |
| Site of pain | 0.51 | Moderate |
| Shifting of pain | 0.52 | Moderate |
| Site of tenderness | 0.57 | Moderate |
| Duration of Pain | 0.83 | Very Good |
| Vomiting | 0.89 | Very Good |
| Temperature | 0.95 | Very Good |

[1] The guidelines for 'interpreting Kappa' are as follows:

| Agreement | Agreement quality |
|---|---|
| < 0.20 | Poor |
| < 0.40 | Fair |
| < 0.60 | Moderate |
| < 0.80 | Good |
| to 1 | Very good. |

**Table 3:** Logistic Regression Analysis for Residents (n = 294 patients)

| Variable | β | std. Error | Wald Statistic | p-value | Critical DEFF |
|---|---|---|---|---|---|
| Age | 0.498 | 0.994 | 0.251 | 0.617 | 0.065 |
| Gender | -0.939 | 0.528 | 3.159 | 0.076 | 0.823 |
| Pain Duration | | | 0.325 | 0.850 | 0.085 |
| Pain Duration (1) | -0.288 | 0.509 | 0.319 | 0.572 | 0.083 |
| Pain Duration (2) | -5.306 | 63.417 | 0.007 | 0.933 | 0.002 |
| Pain Site | | | 0.153 | 0.926 | 0.040 |
| Pain Site(1) | 0.177 | 0.906 | 0.038 | 0.845 | 0.010 |
| Pain Site(2) | 0.440 | 1.124 | 0.153 | 0.696 | 0.040 |
| Pain Type | 0.692 | 0.511 | 1.833 | 0.176 | 0.477 |
| Vomiting | 0.035 | 0.487 | 0.005 | 0.944 | 0.001 |
| Previous Visit | -6.895 | 29.973 | 0.053 | 0.818 | 0.014 |
| Temperature | | | 1.327 | 0.515 | 0.346 |
| Temperature(1) | 0.040 | 0.489 | 0.007 | 0.935 | 0.002 |
| Temperature(2) | -1.911 | 1.695 | 1.271 | 0.260 | 0.331 |
| Tenderness Site | | | 9.971 | 0.007** | 2.597 |
| Tenderness Site(1) | 2.741 | 0.944 | 8.427 | 0.004** | 2.195 |
| Tenderness Site(2) | 0.361 | 1.305 | 0.076 | 0.782 | 0.020 |
| Localized Guarding | 1.863 | 0.508 | 13.469 | 0.000*** | 3.508 |
| Rebound Tenderness | 1.503 | 0.526 | 8.164 | 0.004** | 2.126 |
| Pain Shifting | 0.766 | 0.514 | 2.222 | 0.136 | 0.579 |
| Constant | -5.142 | 1.130 | 20.686 | 0.000 | 5.387 |
| Nagelkerke $R^2$ | 0.568 | | | | |

*p < 0.05
**p < 0.01
***p < 0.001

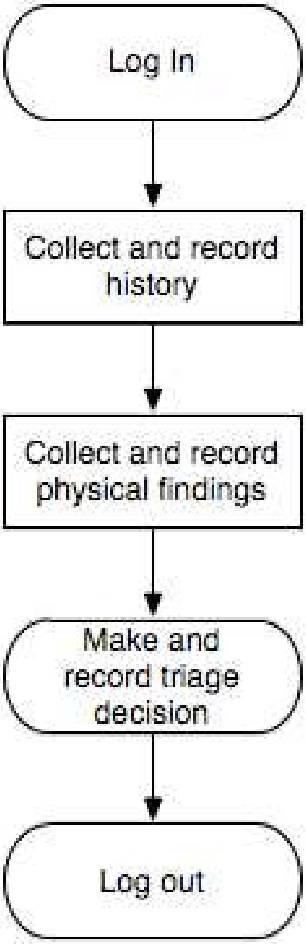
**Table 4:** Logistic Regression Analysis for Staff Physicians (n = 385 patients)

| Variable | β | std. Error | Wald Statistic | p-value | Critical DEFF |
|---|---|---|---|---|---|
| Age | 1.315 | 1.306 | 1.013 | 0.314 | 0.264 |
| Gender | -0.593 | 0.528 | 1.260 | 0.262 | 0.328 |
| Pain Duration | | | 0.614 | 0.736 | 0.160 |
| Pain Duration (1) | 0.377 | 0.514 | 0.537 | 0.464 | 0.140 |
| Pain Duration (2) | -5.517 | 20.305 | 0.074 | 0.786 | 0.019 |
| Pain Site | | | 6.862 | 0.032* | 1.787 |
| Pain Site(1) | 2.467 | 0.973 | 6.429 | 0.011* | 1.674 |
| Pain Site(2) | 2.376 | 1.381 | 2.960 | 0.085 | 0.771 |
| Pain Type | 1.611 | 0.614 | 6.879 | 0.009** | 1.791 |
| Vomiting | 1.299 | 0.601 | 4.674 | 0.031* | 1.217 |
| Previous Visit | 2.691 | 1.417 | 3.604 | 0.058 | 0.939 |
| Temperature | | | 2.312 | 0.315 | 0.602 |
| Temperature(1) | 0.619 | 0.534 | 1.343 | 0.246 | 0.350 |
| Temperature(2) | 2.421 | 2.097 | 1.333 | 0.248 | 0.347 |
| Tenderness Site | | | 3.194 | 0.203 | 0.832 |
| Tenderness Site(1) | 1.082 | 0.953 | 1.288 | 0.256 | 0.335 |
| Tenderness Site(2) | -1.256 | 1.384 | 0.823 | 0.364 | 0.214 |
| Localized Guarding | 1.539 | 0.556 | 7.662 | 0.006** | 1.995 |
| Rebound Tenderness | 2.306 | 0.576 | 16.005 | 0.000*** | 4.168 |
| Pain Shifting | 0.968 | 0.560 | 2.985 | 0.084 | 0.777 |
| Constant | -8.380 | 1.692 | 24.533 | 0.000 | 6.389 |
| Nagelkerke $R^2$ | 0.699 | | | | |

*$p < 0.05$
**$p < 0.01$
***$p < 0.001$

**Figure 1**: MET-AP Screen for Type of Pain


**Figure 2:** MET-AP Usage Scenario


Log In
Collect and record history
Collect and record physical findings
Make and record triage decision
Log out