

**An International Assessment of a Web-based Diagnostic Tool in
Critically Ill Children**

Neal J Thomas MD, MSc, Padmanabhan Ramnarayan MBBS, MRCPCH, Michael J Bell MD, Prabhat Maheshwari MBBS, MD, MRCPCH, Shaun Wilson MRCPCH, Emily B Nazarian MD, Lorri M Phipps RN, MSN, CPNP, David C Stockwell MD, Michael Engel MD, Frank A Maffei MD, Harish G Vyas DM, FRCP, FRCPC, and Joseph Britto MBBS, MD, FRCPC

Pediatric Critical Care Medicine (NJT, LMP) and Public Health Sciences (NJT), Penn State Children's Hospital and The Pennsylvania State University College of Medicine, Hershey, PA, USA; Children's Acute Transport Service (PR), London, UK; Pediatric Critical Care Medicine (MJB,DCS, ME), Children's National Medical Center, Washington DC, USA; Pediatric Intensive Care Unit (PM), St Mary's Hospital, Paddington, London, UK; Pediatric Intensive Care Unit (SW, HGV), Queen's Medical Centre, Nottingham, UK; Pediatric Critical Care Medicine (EBN, FAM), Golisano Children's Hospital, University of Rochester at Strong, Rochester, NY, USA; Isabel Healthcare Inc (JB), Reston, VA, USA

Corresponding Author: Neal J. Thomas, M.D., M.Sc
Pediatric Critical Care Medicine
Penn State Children's Hospital
500 University Drive, MC H085
Hershey, PA 17033 USA

Phone/Fax: (717) 531-5337/ (717) 531-0809

E-mail: nthomas@psu.edu

ABSTRACT

Improving diagnostic accuracy is essential. The extent of diagnostic uncertainty at patient admission is not well described in critically ill children. Therefore, we studied the extent that pediatric trainee diagnostic performance could be improved with the aid of a computerized diagnostic tool. Data regarding patient admissions to five Pediatric Intensive Care Units were collected. Information included patients' clinical details, admitting team's diagnostic workup and discharge diagnosis. An attending physician assessed each case independently and suggested additional diagnostic possibilities. Diagnostic accuracy was calculated using the discharge diagnosis as the gold standard. 206 out of 927 patients (22.2%) admitted to the PICUs did not have an established diagnosis at admission. The trainee teams considered a median of three diagnoses in their workup (IQR 3-5) and made an accurate diagnosis in 89.4% cases (95% CI 84.6%-94.2%). Diagnostic accuracy improved to 92.5% with use of the diagnostic tool alone, and to 95% with the addition of attending physicians' diagnostic suggestions. We conclude that a modest proportion of admissions to these PICUs were characterized by diagnostic uncertainty during initial assessment. Although there was a relatively high accuracy rate of initial assessment in our clinical setting, it was further improved by both the diagnostic tool and the physicians' diagnostic suggestions. It is plausible that the tool's utility would be even greater in clinical settings with less expertise in critical illness assessment, such as community hospitals, or emergency departments of

non-training institutions. The role of diagnostic aids in the care of critically ill children merits further study.

Keywords: diagnosis, diagnostic reminder system, internet, pediatrics

INTRODUCTION

Accurate diagnosis of medical conditions has been of paramount importance for centuries and faulty diagnostic evaluation can lead to patient harm and poor medical management. Three etiologies of diagnostic errors have been recognized within the medical community: no fault errors, system-related errors and cognitive errors . No-fault errors occur either when disease presentation is unusual and unpredictable or when patients are uncooperative. System-related errors occur as a result of either organizational flaws or mechanical problems within the institution. Cognitive errors occur due to lack of adequate medical knowledge or inadequate data collection.

With the explosion of medical literature at exponential rates, avoidance of cognitive errors using emerging technology may be the most efficient method for improving diagnostic skills. A large body of evidence suggests that diagnostic error by clinicians poses a significant problem across a wide variety of settings, comprising 10-30% of adverse medical events (AME) [. One recent observational study suggested that 10% of AME in adult critical care patients were related to misdiagnosis . Moreover, autopsy studies have demonstrated that major diagnoses are missed in critically ill adults and children who die in intensive care units .

We hypothesized that a web-based diagnostic reminder system would produce additional differential diagnosis possibilities for children admitted to five international PICUs who were admitted without an established diagnosis. We further hypothesized that the web-based tool would increase diagnostic accuracy by input of the findings recorded in the medical record admission history and physical examination. To do this, we compared the accuracy of

determining the discharge diagnosis for (i) the admitting team, (ii) an unbiased board-certified/board eligible pediatric intensivist and (iii) ISABEL, a web-based diagnostic reminder system (www.isabelhealthcare.com) .

MATERIALS AND METHODS

This prospective study was conducted during a three-month period at five pediatric intensive care units (PICUs), two from the United Kingdom (UK) and three from the United States (USA). **The selection of ICUs was based on working relationships between the five primary investigators in each site, with a goal of studying institutions that had diverse patient populations, different patient volumes, and a diverse culture.** All patient data collection was approved by the research ethics committees (institutional review boards) at the participating centers prior to initiation of the study. The requirement for informed consent was waived. All admissions to the PICU during the study period were screened by a designated study investigator at each center. Medical admissions to the PICU without an established diagnosis were eligible for study. All surgical admissions and medical admissions with known primary diagnoses (such as culture-proven sepsis, diabetic ketoacidosis and others) were excluded. All PICUs utilize various medical personnel (including residents from pediatrics or pediatric subspecialties, clinical fellows, advanced practice nurses (APN), physician assistants and pediatric intensivists) as their daily clinical team.

Diagnosis by Admitting Team

At each PICU, the screening investigator (pediatric critical care medicine fellow, pediatric resident or pediatric APN) reviewed all admissions and enrolled eligible patients. **Importantly, these investigators neither admitted the child to the PICU nor participated in the clinical care given to the child so that the utility of the tool itself could be tested without bias from difficulties with the system or changes in routine practice.** The medical records of eligible patients were

reviewed and presenting symptoms and physical findings and the differential diagnosis (including all diagnostic possibilities) was extracted. The screening investigator further determined the discharge diagnosis based on review of the medical record at discharge.

Diagnosis by ISABEL

The screening investigator at each center used a customized study version of ISABEL to securely log in, enter, and store data on eligible patients. Each patient was assigned a unique study number, and the date and time of electronic access was logged. Clinical data from the chart **were** entered and was used to drive the diagnostic tool's algorithm for production of diagnostic possibilities.

ISABEL is a Web based diagnostic decision support system (DSS). Computerized decision support tools utilize two or more pieces of clinical data to provide patient-specific recommendations; assist in diverse clinical processes such as diagnosis, prescribing or clinical management; and consist of an underlying knowledge base, user interface for data input, and an inference engine]. The ISABEL knowledge base consists of >100,000 raw text documents describing >10,000 diseases. Documents are drawn from multiple resources such as reputed textbooks (e.g. Oxford Textbook of Medicine) and review articles in journals. The user interface permits data entry in the form of a free text description of clinical findings, without the need for a specific terminology. The inference engine utilizes statistical natural language processing software to search through the entire database of text documents and returns all disease descriptions that match the clinical features entered. This process is similar to a search engine, although much more sensitive and specific. The software indexes

each document in the database daily, and extracts relevant concepts based on frequency of their occurrence, uniqueness and proximity to other concepts. Each diagnosis in the database is assigned clinical weighting scores, derived from expert opinion, to reflect its prevalence by age group (e.g. child), gender and geographical region (e.g. North America). In the current study, when investigators entered historical data and physical findings noted at admission for each patient as search terms, and applied appropriate age, gender and region filters, 10-12 diagnoses that were matched within the database were displayed, arranged by body system (e.g. cardiovascular, respiratory) rather than by clinical probability. Hyperlinks from each diagnostic suggestion led to text documents that matched the search query. There was no limit on the number of search terms that could be entered, although the tool performed well even with 3-5 terms.

A number of different diagnostic DSS have been studied in the past. Tools such as De Dombal's abdominal pain system were developed using data from >6000 patients with proven surgical causes of abdominal pain and their clinical findings. Using probabilistic techniques, this tool assisted clinicians in differentiating surgical from non-surgical causes of abdominal pain []. Other DSS such as Dxplain, Quick Medical Reference, and ILIAD utilized a complex database of relationships between hundreds of diseases and thousands of clinical findings (e.g. rheumatoid arthritis and fever) derived from expert opinion to provide diagnostic hypotheses (i.e. quasi-probabilistic systems) []. Keeping such knowledge bases up to date was a challenge - an expert panel needed to frequently appraise new information, and make significant changes to existing relationships within the database. The ISABEL system utilizes a novel

mechanism to keep its knowledge base current. When updated textbooks or reviews became available on a topic, old documents are simply replaced with new text in the database without any expert input. A small content management team searches monthly for newer editions of textbooks and reviews and updates the database on a quarterly basis. Expert input is only necessary to modify clinical weighting scores, and is undertaken as soon as important new information is available (e.g. SARS virus outbreak in Hong Kong). Since the tool is web based, updates are instantly available to users. A more detailed description of the methodology involved in the development and maintenance of this tool is available].

Diagnosis by Board Certified/Board Eligible Pediatric Intensivist

At each center, a pediatric intensivist independently examined the clinical data available at the time of admission for each patient and the admitting PICU team's diagnostic workup; the discharge diagnosis was withheld from the attending physician during this process. They then highlighted further diagnoses from within the computerized tool's list that they considered clinically relevant or significant for patient assessment. This judgment was based on whether consideration of that particular diagnosis in the workup would have triggered a specific action (either performing further focused clinical examination or a particular test or treatment). If two closely related (or nearly synonymous) diagnoses were displayed in the list, only one was taken into account. This investigator was then able to propose additional diagnoses they felt were relevant to generate an additional list of potential diagnoses.

Diagnostic Accuracy Assessment

Diagnostic accuracy was examined by matching the discharge diagnosis with lists generated by (i) the admitting team, (ii) ISABEL and (iii) the study intensivist. The diagnosis was considered accurate if the discharge diagnosis was located within the list of diagnoses generated by these three methods. An *a priori* calculation demonstrated that 250 patients would be required to identify the frequency of additional significant diagnoses in the diagnostic tool's list assuming 20% of the patients were without an established diagnosis (with 80% power, type I error 5%). Data collection was planned for three months at each unit, based on an estimated annual admission rate of approximately 4000 patients. Diagnostic accuracy rates of admission teams, ISABEL and attending physicians were calculated based on the discharge diagnosis. Results were calculated as proportions, and expressed as percentages with 95% confidence interval limits. The Chi-square test was used to detect statistically significant differences between proportions (defined as p value <0.05). Factors influencing diagnostic accuracy were analyzed by entering the variables into a multiple logistic regression model.

RESULTS

Data **were** collected from all five participating centers during the study. All centers admitted a general mix of medical and surgical pediatric patients. The characteristics of each of the participating centers are shown in **Table 1**. A total of 927 patients were admitted to the five PICUs between May and July 2003. The majority of patients was either surgical/post-surgical or had an established diagnosis by the time they were admitted to the PICU (721/927, 77.8%). A total of 206 patients (22.2%) had no established diagnosis at the time of PICU admission, and were therefore eligible for study; 69 were from UK units and 137 from the US units. Admission data for each center **are** presented in **Table 2**. Of these, 45 lacked discharge diagnoses and were thereby excluded. The remaining 161 patients were therefore analyzed. There were significant differences between the various centers. The number of patients without an established diagnosis at admission varied between center, ranging from 19.6% to 77.8% (mean 38.9%; Chi-square test, $p < 0.05$). The proportion of surgical admissions on the units ranged from 23.7% to 65% (mean 42.9%).

The admitting team's differential list contained a median of three diagnoses per case (IQR 3-5) with a range of 1-12 diagnoses and contained the discharge diagnosis in 144/161 cases (89.4%). In five of the remaining 17 cases in which the admitting team did not include the discharge diagnosis in their initial workup, the diagnostic tool displayed the final diagnosis in the list of suggestions. ISABEL generated a differential list including between 10 and 12 diagnoses and this list contained the final diagnosis in 149/161 cases (92.5%). The attending intensivist accurately provided the final diagnosis in a further four

cases, leading to an overall accuracy rate of 92% (when combined with the admitting team) and 95% (combined with the admitting team and the diagnostic tool). Despite the fact that the diagnostic tool displayed the discharge diagnosis in five patients where the admitting team had failed to consider it, the attending physician was able to identify only two of these as being clinically relevant to the case. The admitting team's diagnostic accuracy appeared to be a function of the age of the patient, but not the number of diagnoses considered at admission or the diagnostic group.

DISCUSSION

In summary, we have demonstrated that within tertiary care PICUs, approximately 20% of critically ill children will be admitted without an established diagnosis and that the clinical team makes an accurate preliminary diagnosis almost 90% of the time. This rate is improved by use of the ISABEL tool as well as a higher degree of clinical training and experience from an intensivist. As we chose to examine every PICU admission during the study period, this tool may have some added utility in a variety of patient populations, particularly those that are complex diagnostic challenges.

In general, it is believed that improvements in technology will ultimately lead to improved outcomes for patients in the health care system of all countries. **While this might be true in some instances, such as the prevention of medication errors]**, there has also been some evidence that increased need for technology **may** also lead to increased patient harm]. To our knowledge, ISABEL is a unique system in that it collates data from multiple sources including textbooks, journal articles and reviews to generate a potential diagnostic list for the user. **However, its ability to improve outcome has yet to be demonstrated, and the direct impact on patient outcome related to use of the system is an area of research that requires further clinical study.** Estimates of diagnostic uncertainty in critical care have come from various sources. In extreme circumstances, persistent diagnostic uncertainty is often only resolved by post mortem studies. A number of autopsy studies have concluded that the rate of clinically significant missed diagnoses in intensive care patients ranges from 2.1% to 21.2%]. In addition, it is reported that in 17% of pediatric autopsies in

critical care, knowledge of the diagnosis pre-mortem would have influenced survival by affecting decision making]. Therefore, it is crucial to suspect and confirm the correct diagnosis early in order to improve outcome in these children. The rate of missed diagnoses directly leading to death is higher in patients with a shorter stay in critical care, while unexpected minor findings are common in patients with a longer ICU stay, presumably due to the addition of new medical problems while undergoing critical care treatments, either nosocomial or iatrogenic]. These post-mortem studies suggest a significant burden of missed diagnoses, and imply that diagnosis constitutes an important part of critical care decision making. We failed to find such a critical burden of missed diagnoses in our study and it did not appear that the missed diagnoses in our patients significantly altered patient care or outcome. However, prospective evaluation of diagnostic errors in the critical care environment suggests that they constitute only a small proportion of medical error; medication-related errors are the most common cause for AME in this setting]. A more rigorous evaluation of the ISABEL tool in a variety of clinical situations may lead to improved knowledge about missed diagnoses as well as a more thorough evaluation of ISABEL's usefulness.

We aimed to explore the role of a diagnostic tool in critical illness as part of this preliminary study. **Recently, ISABEL has been shown to be an effective reminding system in adult emergency departments, with 95% accuracy in presenting the user with the final diagnosis and 90% accuracy in obtaining "must-not-miss" diagnoses]. We undertook the current study in the Pediatric Intensive Care Units of the five centers for several reasons. First, we wanted to demonstrate the utility of ISABEL in a complex population of children with**

complex medical conditions, thereby collecting information on a very challenging population for study. Second, we believe it is important to demonstrate the utility of a tool in diverse patient groups. As three of the PICUs were in the USA and two were in the UK and all units showed similar ISABEL performance, we believe that this highlights the potential utility and generalizability of the tool. Moreover, the patient volumes at the five sites differed by more than 300%, again arguing for the utility of the tool in a diverse population.

The most interesting question that can be asked regarding our data and device is whether a clinician can improve his/her own diagnostic ability in collaboration with the use of ISABEL. It is not intended that ISABEL is to become a substitute for clinical acumen, but rather an ancillary tool to establish potential diagnoses. We believe that this is consistent with others who believe that the performance of the intended user *working in combination with* the tool is greater than the performance of the user alone]. Obviously, this necessitates that the user-computer interaction must be seamless and user-friendly. None of the investigators in our subset of more than a dozen had difficulty in utilizing the ISABEL tool to generate useful diagnostic lists. Again, further studies with more diverse clinician populations may also demonstrate whether this interface is sufficient for all users' needs.

This study has a number of limitations. Patient selection of only cases without an established diagnosis at admission to ICU may not accurately indicate the requirement or need for diagnostic decision making. Since diagnoses are being constantly made and adjusted within the first hours of PICU admission, it is possible that ISABEL might have other uses than the method we tested. Our methods also did not allow for testing in other settings, such as the emergency

department, where less data are available and diagnostic uncertainty might be considerably greater. Second, we only studied this tool in tertiary care facilities with well-trained clinicians and highly-productive trainees. The results of a similar study to this in a smaller unit are unpredictable at this time. It is possible that the experience of the trainees masked the true utility of the ISABEL device. Conversely, it is possible that in a smaller unit the diagnoses might be less challenging and this might make ISABEL's utility even less. Only future studies can answer this question. Lastly, we intentionally excluded surgical cases in our study population even though they made up almost over 50% of total admissions to the PICUs. Therefore, it is not possible to assess the utility of ISABEL in diagnosing common pediatric surgical emergencies such as ruptured appendix or other disorders.

In conclusion, ISABEL improved the diagnostic accuracy of the clinical team caring for children without an established diagnosis upon PICU arrival. Future studies demonstrating ISABEL's utility in other patient populations will be required to fully assess whether web-based tools can improve children's outcome after critical illness. **It will be important to determine if implementation of tools such as ISABEL can be seamlessly integrated within clinical practice and if subsequent versions of the tool can be sufficiently rigorous to advance as the field of pediatrics evolves and changes. This challenge must be met by constant update and maintenance of diagnostic tools, as well as ongoing evaluation of the efficacy of these tools in practice.**

ACKNOWLEDGEMENTS

The computerized diagnostic tool studied was provided free for registered users by the Isabel Medical Charity at the time of this study. It is now managed by a commercial entity called Isabel Healthcare and is available only to subscribers. Dr P Ramnarayan currently advises Isabel Healthcare on research activities on a part-time basis. Dr J Britto is Clinical Director of Isabel Healthcare Inc, USA. All the other authors have no competing interests.

REFERENCES

Table 1: Key baseline characteristics of the participating critical care

units. * Total annual admissions at all participating centers: 3794. (C = center; UK= United Kingdom; US= United States)

	C1	C2	C3	C4	C5
Country	UK	UK	USA	USA	USA
Total number of beds	6	10	12	12	26
Medical admissions in 2003 (%)	143 (35)	363 (77)	289 (37)	402 (62)	1001 (67)
Surgical admissions in 2003 (%)	267 (65)	106 (23)	483 (63)	248 (38)	492 (33)
Total annual admissions*	410	469	772	650	1493

Table 2: Breakdown of patient admissions according to diagnostic groups: May-July 2003. * Difference between centers was statistically significant (chi-square test, $p < 0.05$). (C = center)

	Known surgical diagnosis (%)	Known medical diagnosis (%)	Unknown diagnosis (%)*	Total
C1	67 (65.0)	8 (7.8)	28 (27.2)	103
C2	22 (22.7)	33 (34)	42 (43.3)	97
C3	117 (61.3)	30 (15.7)	44 (23)	191
C4	60 (38.2)	78 (49.7)	19 (12.1)	157
C5	132 (34.8)	174 (45.9)	73 (19.2)	379
Total	398 (42.9)	323 (34.8)	206 (22.2)	927