ORIGINAL ARTICLE

ISABEL: a web-based differential diagnostic aid for paediatrics: results from an initial performance evaluation

P Ramnarayan, A Tomlinson, A Rao, M Coren, A Winrow, J Britto

.....

Arch Dis Child 2003;88:408-413

Aims: To test the clinical accuracy of a web based differential diagnostic tool (ISABEL) for a set of case histories collected during a two stage evaluation.

See end of article for authors' affiliations

Correspondence to: Dr J Britto, Department of Paediatrics, Imperial College School of Medicine, St Mary's Hospital, Norfolk Place, London W2 1PG, UK; j.britto@ic.ac.uk

Accepted 4 October 2002

Methods: Setting: acute paediatric units in two teaching and two district general hospitals in the southeast of England. Materials: sets of summary clinical features from both stages, and the diagnoses expected for these features from stage I (hypothetical cases provided by participating clinicians in August 2000) and final diagnoses for cases in stage II (children presenting to participating acute paediatric units between October and December 2000). Main outcome measure: presence of the expected or final diagnosis in the ISABEL output list.

Results: A total of 99 hypothetical cases from stage I and 100 real life cases from stage II were included in the study. Cases from stage II covered a range of paediatric specialties (n = 14) and final diagnoses (n = 55). ISABEL displayed the diagnosis expected by the clinician in 90/99 hypothetical cases (91%). In stage II evaluation, ISABEL displayed the final diagnosis in 83/87 real cases (95%).

Conclusion: ISABEL showed acceptable clinical accuracy in producing the final diagnosis for a variety of real as well as hypothetical case scenarios.

large proportion of routine clinical decision making depends on the availability of good quality clinical information and instant access to up to date medical knowledge.¹ There is increasing evidence that the use of computer aided clinical decision support to manage medical knowledge results in better healthcare processes and patient outcomes.² However, the use of computer based knowledge management techniques in medical decision making has remained poor.³⁴

Considerable advances have been made in making latest processed information from clinical trials available, exemplified by the Cochrane database and the Clinical Evidence series.⁵ The free accessibility of Medline on the internet (http://www.ncbi.nlm.nih.gov/PubMed) has enabled easy and universal search of the medical literature. However, these endeavours do not primarily serve the busy clinician at the bedside seeking bottom line answers to routine clinical questions. These questions involve diagnosis and immediate management for a patient in general practice⁶ or in the emergency department.⁷ Recent initiatives such as the ATTRACT project have attempted to provide quick, up to date answers to clinical queries for the bedside physician, but involve a substantial financial and resource commitment.⁸

In addition, the current model of healthcare delivery within the UK National Health Service (NHS) accentuates these problems by providing care in the form of an "inverted pyramid of knowledge". Clinical wisdom and knowledge are concentrated at the top among senior staff, in many cases distant from the patient who is first seen by a junior doctor at the bottom of the pyramid. This may contribute to a significant proportion of medical error,⁹ resulting in extra bed days and a preventable financial burden.¹⁰

ISABEL (Isabel Medical Charity, UK) is a computerised differential diagnostic aid for paediatrics that is delivered via the world wide web. It was developed following a missed diagnosis on a 3 year old girl with necrotising fasciitis complicating chicken pox. It has currently over 9000 registered users (including doctors, nurses, and other healthcare professionals) and receives over 100 000 page requests every month. Powered by Autonomy, proprietary software that serves as an efficient information retrieval engine by matching patterns within unformatted text, the tool produces differential diagnoses for any set of clinical features by searching text from standard paediatric textbooks. Rather than provide a single diagnosis (a diagnostic tool), ISABEL is primarily intended to suggest only a differential diagnosis and serve as a "reminder" system to remind the clinician of potentially important diagnoses that might have been missed.

During the development of ISABEL, text from each textbook pertaining to each of 3500 different diagnostic labels (for example, measles, migraine) was added to the ISABEL database. These textbooks included *Nelson's textbook of pediatrics* (16th edition, 2000, WB Saunders), *Forfar and Arneil's textbook of paediatrics* (5th edition, 1998, Churchill Livingstone, UK), *Jones and Dargan Churchill's pocket book of toxicology* (2001, Churchill Livingstone, UK), and *Rennie and Roberton's textbook of neonatology* (3rd edition, 1999, Churchill Livingstone, UK). Each diagnostic label was allocated an age group classification ("newborn", "infant", "child", or "adolescent") to prevent inappropriate diagnoses being presented to the user for a specific patient age group.

In order to examine ISABEL's utility, an evaluation programme was planned in a stepwise fashion: initial system performance to establish the safety of the tool, subsequent evaluation of impact in a simulated setting, and evaluation of impact in a real life clinical setting. This paper describes only the initial evaluation of ISABEL's capability, and focuses on system performance. The tool was isolated from intended users (clinicians); its impact on clinical practice and decision making was not examined. This study was planned in two stages in two different settings, with the following considerations:

- The two stages to be conducted in series for ease of data collection.
- They were intended to provide a variety of hypothetical as well as real cases for the investigators to test the tool.



• ISABEL is useful primarily as a reminder tool and not as an "oracle". However, in reminding clinicians of diagnoses in a given clinical setting, it is imperative that the final diagnosis is also *one* of the "reminders" suggested. ISABEL would be considered "unsafe" for clinical use if many plausible diagnoses were suggested to the junior doctor, but the final diagnosis was not. For this reason, it is crucial that ISABEL showed acceptable clinical accuracy by displaying the final diagnosis (especially for inexperienced junior doctors who may not have considered the "correct" diagnosis).

METHODS

Differential diagnostic tool

ISABEL was delivered on the internet free of charge (www.isabel.org.uk). During the development phase, only the investigators had access to ISABEL, enabled by a secure log-in procedure. This would ensure that clinicians participating in the data collection would not be able to use ISABEL and impact on clinical management. On accessing the tool, the patient age group had to be chosen first (newborn, infant, child, or adolescent). Following this, the summary clinical features of the case were entered into a free text box. These features would normally be gathered from the history, physical examination, and results of initial investigations. Any additional findings could be subsequently entered into the tool to focus the differential diagnosis further.

To maximise the tool's performance, clinical features had to be entered in appropriate medical terminology (rather than lay terms, as the database consisted of textbooks), the spelling had to be accurate (British or American), and laboratory results needed interpretation in words ("leucocytosis" or "increased white cell count" for a white cell count of 36.7×10^6 /µl). These data were then submitted to the ISABEL database, searched by Autonomy, and a fresh web page displaying the results from ISABEL was produced. This list included around 10–15 unique diagnoses for consideration, classified on the basis of the systems from which they were drawn (respiratory, metabolic, etc). The diagnoses were not ranked in order of probability, thus reinforcing the function of the tool as a "reminder" system.

Study design and conduct

Stage I: We undertook this study in August 2000 within the Department of Paediatrics at St Mary's Hospital, London, which has both general paediatric as well as specialist services in infectious diseases, neonatology, and intensive care. Clinicians with varying levels of experience (consultant, registrar, and senior house officer) were contacted to provide hypothetical case histories of acute paediatric presentations. For each case, they specified the age group category, summarised the clinical features, and listed the expected diagnosis(es).

Stage II: This study was undertaken in four acute paediatric units, two teaching hospitals (St Mary's Hospital, London and Addenbrookes' Hospital, Cambridge), and two large district general hospitals (Kingston Hospital, Surrey and Royal Alexandra Hospital, Brighton). This study was done from October to December 2000. Junior doctors working within these departments prospectively collected data on children presenting to the acute paediatric unit. Only data regarding the age group, a summary of clinical features at initial presentation, and the working diagnosis (es) were collected from the doctors. The final diagnosis for each patient, as decided by the clinical team at the end of the hospital stay (or at the end of the clinical assessment), was collected from the discharge summary.

Guidance was provided to the doctors to specify clinical features in medical terminology and to interpret results of initial investigations in words. This was done so that the investigators would not need to modify the content provided before entering the clinical features into ISABEL. One investigator (AT) collected these data in a structured form. In one sitting at the end of the data collection, she then entered the age group and the clinical features of each case into ISABEL as provided by the junior doctors, without modifying the content or spelling. This preserved the inherent user variability in summarising clinical features. Results generated by the tool for each case were collected for analysis.

Outcome measures

The main outcome measure used for both stages of the study was the presence of the expected or final diagnosis(es) within the results generated by ISABEL. This was a measure of the



Figure 2 Sample screen from ISABEL: input of clinical features.

and Isabel	TOTAL CONVENTION AND A DESCRIPTION	Deports for Debut Aprilia
		and the second second second space second se
** 1++++1 *** ** Diagnostic Tool:		These Thesises Terms & sendites These Testers
Have you considered?		
Infectious Discases	Respiratory System	
Respiratory Syncytial Virus / Bronchiolitis Pneumonia Pertussis	Cystic Fibrosis Upper Respiratory Infection (Nasopharyngitis) Congenital Lober Emphyseme	
Allergic Disorders	Bronchial Foreign Bodies Dulmonary Henoplasia	
Asthma	Pulmonary Oedema	
Cardiovascular System		
Heart Failure		
and an and the second se		
Would you like to contribute a clinical	experience related to this situation?	
Would you like to contribute a clinical Search again:	experience related to this situation?	
Would you like to contribute a clinical Search again: Patient is assessing with the	experience related to this situation?	
Would you like to contribute a clinical Search again: Patient is intercent of which the Coup Wearch Increased work of breaking Macount conversa Where and cruckies on comparison Therbypone Fuer Freeding	following clinical features:	
Would you like to contribute a clinical Search again: Patient is interced days-iyees with the Cough Whether and work of interthing Intercent work of interthing Intercent contaction Therbysone Techysone Techysone Pour Freeding	following clinical features:	
Would you like to contribute a clinical Search again: Patient is interplatery with the Crooph Weaton Therement work of breaking management contention Techypoon and cracking on semication Techypoon	following clinical features:	

Figure 3 Sample screen from ISABEL: output of diagnostic reminders classified by system.

Primary diagnosis expected by paediatrician ar	nd num	ber of forms	
Acute graft versus host disease	1	Juvenile psoriatic arthritis	1
Acute pyelonephritis	1	Juvenile spondyloarthropathy	1
AIDS	1	Kawasaki disease	3
Aspergillosis	1	Leptospirosis	1
Bacterial meningitis	1	Lyme disease	2
Bacterial pneumonia	1	Malaria	7
Blastomycosis	1	Measles	1
Brucellosis	1	Meconium aspiration syndrome	1
Campylobacter	1	Meningitis	2
Cat scratch disease	1	Meningococcal disease	3
Cellulitis	1	Meningococcus	1
Chlamydia pneumoniae infection	1	Mumps	2
Chlamydia trachomatis infection	1	Mycoplasma infection	1
Chronic granulomatous disease	1	Necrotising enterocolitis	1
Congenital cytomegalovirus infection	1	Nonpolio enterovirus infection	1
Crohn's disease	1	Osteomyelitis	1
Croup	3	Parotitis	1
Dermatomyositis	1	Pauciarticular juvenile chronic arthritis	1
<i>E coli</i> 0157 infection and haemolytic uraemic	1	Periorbital cellulitis	1
syndrome			
Epstein-Barr virus infection	1	Pneumonia	1
Endocarditis	1	Polio	1
Endotracheal tube obstruction	1	Rheumatic fever	1
Enteric fever	1	Respiratory syncytial virus bronchiolitis	1
Enteroviral infection	1	Respiratory syncytial virus infection	1
Erythema infectiosum	1	Scarlet fever	1
Gastroenteritis	1	Sepsis	1
Gram negative septicaemia	1	Shiaella infection	1
Haemolytic disease of newborn	1	Staphylococcal scalded skin syndrome	1
Hand, foot, and mouth disease	1	Staphylococcal toxic shock	1
Herpes simplex gingivostomatitis	1	Still's disease (systemic juvenile chronic arthritis)	1
Herpes zoster (shingles)	1	Subacute sclerosing panencephalitis	1
HIV	2	Systemic sclerosis	1
Human herpes virus infection	1	Tuberculous pneumonia	1
latrogenic blood loss	1	Toxic shock syndrome	4
Impetiao	1	Tuberculosis	1
Infant botulism	2	Uringry tract infection	2
Infant of a diabetic mother	1	Varicella zoster (chicken pox)	1
Infection mononucleosis (alandular fever)	i	Viral encephalitis	1
Infectious mononucleosis	1	Viral uringny respiratory tract infection	1

clinical accuracy of the tool, defined as reminding clinicians of the "correct" diagnosis.

RESULTS

Figure 1 summarises the outcome of all data that were collected for both stages of the evaluation. On the whole, 99 hypothetical cases (provided by 13 clinicians) were eligible for analysis in stage I, and 100 cases in stage II. Forms in which no diagnoses were entered and where the clinical features section included the final diagnosis in the wording were excluded from testing. Figures 2 and 3 show sample screens from ISABEL for a real case in stage II testing.

Tables 1 and 2 summarise the clinical characteristics of cases collected in stages I and II, providing an indication of the spectrum and frequency of final diagnoses as well as special-ties covered during the evaluation.

Stage I: Presence of the expected diagnosis/es in the ISABEL differential diagnosis list: Of the 99 hypothetical cases that were used to test ISABEL, the expected diagnosis (if it was a single diagnosis) and *all* the expected diagnoses (if the clinician expected to see multiple diagnoses) were present in 90 cases (91%).

Stage II: Presence of the final diagnosis in the ISABEL differential diagnosis list: One hundred real cases were used to test ISABEL. Each of these cases had a primary final diagnosis which was used as the outcome variable (11 cases also had a supplementary diagnosis, which was not used to test the tool). Table 3 shows an example from stage II evaluation data.

In 13/100 cases, the final diagnosis was non-specific (such as "viral illness"). Since textbooks do not describe such nonspecific diagnoses, and the ISABEL database was itself created from textbooks, these were not recognised as distinct final diagnoses. In the remaining 87 cases, the final diagnosis was present in the ISABEL output in 83 cases (95%). The four cases in which the final diagnosis was absent in the ISABEL output were: Stevens-Johnson syndrome, respiratory syncitial virus bronchiolitis, erythema multiforme, and staphylococcal cellulitis.

Recognising that non-specific final diagnoses are often made in routine clinical practice, ISABEL was separately tested against these diagnoses. In 10/13 such cases, ISABEL suggested diagnoses that were nearly synonymous. For a nonspecific final diagnosis such as "viral illness", ISABEL suggested alternatives such as roseola infantum, Epstein-Barr virus, and enteroviruses.

For each case, the ISABEL differential diagnostic tool suggested more than 10 diagnoses (mode 13, range 10–15).

DISCUSSION

Presence of the expected/final diagnosis

This outcome measure is akin to the "sensitivity" of a test. In this respect, ISABEL showed a level of sensitivity between 83% and 91%. More importantly, ISABEL displayed the final diagnosis in stage II cases even when only the presenting clinical features were entered. Although the primary function of the tool is to remind clinicians to consider *other* reasonable

System	Final diagnosis (n=55)	Primary diagnosis*	Supplementary diagnosis*
Allergy	Allergic reaction	2	
	Egg allergy	1	
	Urticarial reaction		1
Dermatoloav	Erythema multiforme	2	
5	Staphylococcal infection		1
Endocrine	Diabetes	1	
	Diabetic ketoacidosis	2	
Gastroenterology	Foreign body ingestion	1	
	Non-specific aastritis	1	
	Non-specific abdominal pain	2	
	Pyloric stenosis	1	
Haematoloav	Sickle cell crisis	2	
Infection	Acute aastroenteritis	-	1
Infection	Cellulitis	1	
	Cervical lymphadenitis	1	
	Enteroviral infection	2	
	Infected insect bite	1	
	Mycoplasma infection		1
	Osteomyelitis	1	
	Otitis media	2	1
	Pneumococcal meningitis	2	1
	Proumonia	7	
	Proguricular abscoss	1	
	Presental collulitie	2	
	Posoola infantum	1	
	Potavirus infaction	1	
	Puballa	1	
	Socied de la sundromo	1	
	Scalded skill syndrome	1	
	Stevens Johnson aundreme	1	1
	Stevens-jonnson syndrome	1	1
		1	I
		2	
	loxic snock syndrome	1	
		2	
	Viral mess	9	
Nanata		1	
	Opidie willidrawal	1	
пертгоюду	I lain and the sting of a sting	1	
NI 1		1	
Neurology		1	
Description	Vasovagal affacks	1	
Respiratory	Acute severe asthma	I	1
		<i>r</i>	1
	Asthma	5	I
	Bronchiolitis	5	
	Croup	2	
		10	
	Kespiratory syncytial virus bronchiolitis	12	1
	A subsequences where the second	5	1
Kneumatic	Acute exacerbation of systemic lupus	I	
	erythematosus	0	
	nenoch-schonlein purpura	2	
	Nawasaki disease	1	
Skeletal	Diskifis of LT/LZ	1	
с : I	Supped upper temoral epiphysis	1	
surgical	Appendicitis	1	

diagnoses in their diagnostic plan and thus influence their management plan, it is important that the tool also generates accurate diagnoses in its output. The use of a final diagnosis as the gold standard is useful in this regard, to ensure that even inexperienced clinicians using the system remain "safe". In this context, diagnostic accuracy rates of clinicians, in an unselected patient population, have been around 60%.¹¹ These studies used necropsy findings or results of specific "diagnostic" tests as the gold standard. In a prospective evaluation of a prototype of the medical diagnostic aid Quick Medical Reference (QMR), the unaided diagnostic accuracy rate of physicians (an entire ward team considered as a single unit) was 60% for diagnostically challenging cases.¹² In this highly

selective patient population, it was possible to establish a final

diagnosis only in 20 out of 31 cases, even after follow up for six months and extensive investigation.

In general, published reports of similar evaluations of other diagnostic systems are sparse. In one study, some commonly used systems for adult medicine (Dxplain, Iliad, QMR, and Meditel) had an accuracy of between 50% and 70% when tested against a set of challenging cases.¹³ Our figures compare favourably with results obtained from testing other diagnostic systems in clinical use.^{14–16} Most such systems are expert systems, and use a combination of rule based and Bayesian approaches to model diagnostic decision making.¹⁷ Most systems are also oriented towards adult medicine. Even the few existing paediatric diagnostic aids offer support related to very specific areas such as rheumatic diseases and abdominal

Table 3	Example from s	stage II evaluation data
showing	the clinical featu	res, the ISABEL differential
diagnos	tis list, and the fine	al diagnosis
Age group	Clinical features	Final diagnosis

nfant	Cough	Respiratory syncytial virus bronchiolitis
	Wheeze	
	Increased work of	
	breathing	
	Subcostal recession	
	Wheeze and crackles or examination Tachypnoea	
	Poor feeding	

pain, making them difficult to use in routine clinical practice.¹⁶ ¹⁸ The evaluation of one such system for paediatric rheumatic diseases suggested 80% accuracy (92% when only diagnoses included in its knowledge base were used).¹⁶

Limitations of the study

This study was designed to evaluate only the presence of the final diagnosis in the reminder list. It did not address the plausibility of the other diagnoses suggested by ISABEL. In this respect, it is possible that although the final diagnosis was present in the differential diagnosis list, the other suggestions made by ISABEL were misleading. However, in order to verify the safety of the tool, it was crucial to show that the final diagnosis formed part of the "reminder" list. Further evaluation is underway to test the relevance of the alternate diagnoses.

The hypothetical cases did not represent the complete spectrum of paediatrics and may not have tested the true limits of the tool. Similarly, the number of real cases used to test the tool may have been insufficient to ensure a complete evaluation of the system's capability, especially in the specialities of neonatal paediatrics, oncology, and paediatric surgery. In addition, this study was conducted in secondary hospital care. Further testing is planned in a primary care setting to assess the safety of the tool.

Considering that this study was only a preliminary evaluation of the performance of the tool, an attempt to extrapolate these results to routine clinical use is not easily possible. This evaluation separates the tool from the intended user (the clinician). This tool, like other similar tools, will only be used as an adjunct to the existing system of clinical decision making¹⁹ Despite this, it is possible that inexperienced doctors using the tool either in a primary care or hospital setting might cause an increase in referrals and investigations. The true measure of the tool's benefits and risks will be clear only when tested by real clinicians uninvolved in the development of the tool. These questions will be answered in subsequent clinical impact evaluations.

CONCLUSIONS

Our study shows that in a large proportion of cases, ISABEL has the potential to remind the clinician of the final diagnosis in a variety of hypothetical as well as real clinical situations.

The spectrum of cases tested is broad enough to ensure that the tool is not limited to a special subset of users. Delivered via the world wide web, ISABEL could impact on patient management on a global level. Because of a potential synergy between the doctor and ISABEL, the results of this study cannot be directly extrapolated to estimate the clinical impact of the tool. Further studies are underway to evaluate the effects of clinicians using ISABEL to aid clinical decision making in simulated as well as real settings. These studies are essential to explore the positive as well as negative impact ISABEL might produce on healthcare processes, health economics, as well as patient outcomes.

Authors' affiliations

P Ramnarayan, A Tomlinson, M Coren, J Britto, Department of Paediatrics, Imperial College School of Medicine, St Mary's Hospital, Norfolk Place, London, UK

A Rao, Department of Haematology, Chase Farm Hospital, The Ridgeway, Enfield, UK

A Winrow, Department of Paediatrics, Kingstone Hospital, Galsworthy Road, Kingston upon Thames, Surrey, UK

REFERENCES

- Wyatt JC. Clinical knowledge and practice in the information age: a handbook for health professionals. London: The Royal Society o Medicine Press, 2001
- 2 Hunt DL, Haynes RB, Hanna SE, et al. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA 1998;280:1339-46.
- 3 Ely JW, Osheroff JA, Ebell MH, et al. Analysis of questions asked by
- family doctors regarding patient care. BNJ 1999;319:358-61 4 Hersh WR, Hickam DH. How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review. JAMA 1998;280:1347-52.
- 5 Godlee F, Smith R, Goldmann D. Clinical evidence. BMJ 1999;318:1570-1
- 6 Gorman PN, Ash J, Wykoff L. Can primary care physicians' questions be answered using the medical journal literature? Bull Med Libr Assoc 1994:82:140-6.
- 7 Wyer PC, Rowe BH, Guyatt GH, et al. Evidence-based emergency medicine. The clinician and the medical literature: when can we take a shortcut? Ann Emerg Med 2000;36:149-55.
- 8 Brassey J, Elwyn G, Price C, et al. Just in time information for clinicians: a questionnaire evaluation of the ATTRACT project. BMJ 2001;322:529-30.
- Neale G, Woloshynowych M, Vincent C. Exploring the causes of
- adverse events in NHS hospital practice. J R Soc Med 2001;94:322-30. 10 Alberti KG. Medical errors: a common problem. BMJ 2001;322:501-2.
- Sadegh-Zadeh K. Fundamentals of clinical methodology. 4. Diagnosis. Artif Intell Med 2000;20:227-41.
- 12 Bankowitz RA, McNeil MA, Challinor SM, et al. A computer-assisted medical diagnostic consultation service. Implementation and prospective evaluation of a prototype. Ann Intern Med 1989;**110**:824–32
- 13 Berner ES, Webster GD, Shugerman AA, et al. Performance of four computer-based diagnostic systems. N Engl J Med 1994;330:1792-6.
- 14 Feldman M, Barnett GO. Pediatric computer-based diagnostic decision support. Scientific Program, Section on computers and other technologies. American Academy of Pediatrics, New Orleans, LA, 27 October 1991.
- 15 Swender PT, Tunnessen WW Jr, Oski FA. Computer-assisted diagnosis. Am J Dis Child 1974;127:859-61
- 16 Athreya BH, Cheh ML, Kingsland LC. Computer-assisted diagnosis of pediatric rheumatic diseases. Pediatrics 1998;102.
- 17 Shortliffe EH, Perreault LE. Medical informatics. New York:
- Springer-Verlag, 2001. 18 **De Dombal FT**, Leaper DJ, Staniland JR, et al. Computer-aided diagnosis of acute abdominal pain. BMJ 1972;2:9-13.
- 19 Ramnarayan P, Britto J. Paediatric clinical decision support systems Arch Dis Child 2002;87:361-2.