CME Article

Critically appraised topics and evidence-based medicine journals

M Dawes

INTRODUCTION

Medical knowledge has expanded rapidly over the last 50 years. The number of subcategories of disease, the number of diagnostic tests and the number of therapies have all increased dramatically while the prognosis of many conditions continues to improve. Simultaneously, the volume of scientific articles published is doubling every 10 years[1]. It is against this increasing volume of information and increasing complexity of health care that evidence-based practice was developed to help health practitioners[2].

THE VOLUME OF QUESTIONS

The building blocks of the science of medicine are a series of answers to questions. The taxonomy of these questions is familiar to every medical student and includes aetiology, prevalence, incidence, diagnostic tests and therapy. As we go through medical school, we learn the answers to many of those questions. Sadly, a lot of those answers are out-of-date, or proven wrong, by the time we finish our postgraduate training. But worse than that is we do not necessarily know which ones are wrong or out-of-date.

Family physicians see people who may have any clinical condition, and yet are expected to know everything about those conditions. If you are looking after an average list of patients as a general practitioner, you will be making some 2,500 new diagnoses each year. This will cover approximately 500 individual diseases[3]. For each disease, you need to know approximately 34 answers to the basic groups of clinical questions (Table I), although this will clearly vary depending on the disease or problem.

The answers to these questions differ with age, gender and ethnic group of the patient. We need answers for children, young adults, middle-aged and the elderly. This increases the numbers of questions and answers in each of the major question groups (except perhaps aetiology) by a factor of 20 to account for these variables, so bringing the numbers of questions per disease to 660. That means, conservatively, a potential total of 333,000 questions that a generalist may need answers for to practise safely and effectively. As one looks down the list and thinks about how much information we have to know, one can immediately see how difficult the task is for general practitioners. For example. you may have a patient with arthritis who is 56 years old and wants to know whether she should take glucosamine. You then have to find out the outcome she wants from therapy. Is it reduction of pain, living independently, increased mobility, or all three? You then have to know how likely she would gain benefit from therapy in terms of any of these outcomes. You have to know the side effects and possible interactions, and you have to have all this available as you see the patient.

Medicine has become much more complex in the last 50 years, with the volume of world literature expanding rapidly[1]. There were 27 randomised control trials published last year just about otitis media. There is no possibility that by reading a couple of journals per week and attending postgraduate seminars, that doctors will be able to keep up-to-date with everything that is important that may affect their patients. At the same time, we need to make sure we are doing the right things to the right patients at the right time[4]. This task may seem impossible for the generalist.

Table I. Grouping of clinical questions per disease.

Major question groups	Number of questions per disease
Aetiology	1
Prevalence	1
Incidence	1
Diagnostic tests (5 per disease)	5
Therapies (5 per disease)	
• Benefit (3 different outcomes per therapy)	15
• Side effects	5
• Drug interactions	5
Prognosis	1
Total	34

Department of Family Medicine McGill University 515-517 Pine Avenue West Montreal, Quebec Canada H3P 2S3

M Dawes, MD, MBBS, FRCGP
Chair

Correspondence to:
Dr Martin Dawes
Tel: (514) 398 7375
Fax: (514) 398 4202
Email: martin.dawes@mcgill.ca

Technology has also developed at the same time as the information explosion. The computer should have made life easier for clinicians but paradoxically at present it is more difficult. The computer has allowed us to store information over the last 30 years but without developing very effective retrieval systems. Simultaneously, we have begun to realise that not all research have the same level of validity. As a result we are left with two major problems. Firstly, it is difficult to find the "right" information. As anyone who has tried to search Google Scholar (http://scholar.google.com) or MEDLINE quickly realises, "simple, quick and effective searches" do not exist. Secondly, we do not know whether to believe the information we do find.

THE NEED FOR APPRAISAL

To answer any question, we can go to experts, textbooks, the world literature, or appraised secondary sources of information. Experts can give very good advice but they tend to have a biased view of the disease. They see patients who frequently are a lot sicker than those we may see in family medicine and so their approach in terms of diagnosis or therapy is often, quite appropriately, much more aggressive. Whereas in family medicine, we frequently want to "rule out" diagnosis in a well patient; in specialty medicine, the need is to "rule in" a diagnosis in a critically-ill patient. This is an oversimplification of the difference between community and hospital practice but the prevalence of disease is much higher in the latter and this does lead to a different style of practice.

Textbooks are also extremely useful for getting "background" information about any topic but they are often several years out-of-date when they are published. The world literature of published research in the field of medicine is vast. It is kept on many databases, the most widely known being MEDLINE. However, there are many other sources of medical literature. There is no hard data on the number of medical journals in existence around the world but it may be in the region of 15,000 and MEDLINE only takes abstracts from a third of these. This means it may be biased towards articles with positive results and we have to be careful when just using one database in assuming all the research has been found.

Secondly, the information contained in these primary research journals has not been appraised. The need to check for validity seems on the surface to be bizarre. Surely this is what journals should be doing. That assumption is correct, but in reality, the job of a publisher is primarily to make a profit for its shareholders. Therefore, one will find articles with flawed trials in even the most highly-cited journals. It is not enough to accept at face value that a trial is believable because it is in a "reputable" journal. There is now overwhelming evidence that trial methodology has a major influence on the results of therapeutic trials[5,6].

Concealment of randomisation, masking and randomisation itself are three main components of the randomised trial design. Concealment of randomisation occurs when the recruiter of the patient to the trial is unaware of what the patient will receive if they enter the trial. If there is no concealment of randomisation, this may exaggerate the efficacy of the treatment by as much as 30% more than trials where there is adequate concealment. If there is no randomisation, then the effect size may be exaggerated by 20%. Surprisingly perhaps, of lesser importance is blinding (masking) – the doctor looking after the patient, or the patient themselves, knowing whether they are giving (being given) the experimental or placebo treatment. If there is no blinding, the results may exaggerate the effectiveness of the treatment by 15%. If less than 80% of patients are followed-up, one cannot tell what is happening and the results become meaningless. Problems with trial design are not just related to studies on therapy, but also apply to diagnostic studies[7] and systematic reviews[6,8]. Despite the QUOROM statement[9] setting out how systematic reviews should be presented, their quality may still be poor[8].

Journals started as the diaries of researchers. These were then passed around among colleagues so that the information could be shared. The development of trials and the assessment of bias are subsequent to that descriptive era. What we are now faced with are seemingly complex long papers with obscure statistical analyses. The scenario is no different to a medical student facing their first anatomy class. By reading around the subject and by practice, one is able to quickly identify key anatomical structures. The same is true of reading scientific articles. One can learn certain aspects of trial design very quickly and be able to assess whether an appropriate design was used for a certain research question.

This does mean finding the correct article, having the full text in front of you and then the time, skills, and confidence to be able to undertake that. Therefore, this process is not possible at the point of clinical care. Secondary sources of pre-appraised evidence, such as the journal, Evidence-Based Medicine, critically appraise the literature from over 100 journals. This process highlights any potential bias for the reader.

Fig. 1 An example of a CAT.

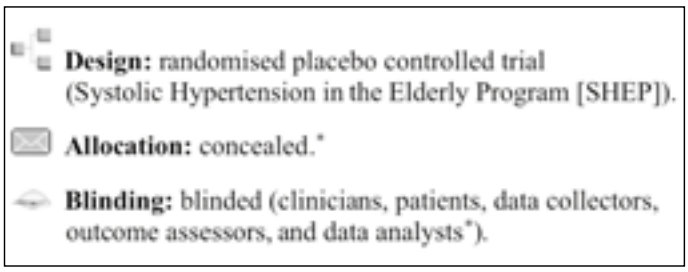

Fig. 2 Use of graphical icons to denote various elements of trial design in a CAT[22].

THE PROCESS OF APPRAISAL

Appraisal of health care studies has been addressed in this series[10]. However, it is often useful to have several alternative texts to help you understand this process. These three textbooks are examples that have come from the development of workshops or courses and are a useful foundation to anyone wanting to know more about appraisal:

1. Users' Guides to the Medical Literature: A Manual for Evidence-based Clinical Practice[11].
2. Evidence-Based Medicine: How to Practice and Teach EBM[12].
3. Evidence-Based Practice. A Primer for Health Care Professionals[13].

To assess a study, there are checklists (some of which have now been incorporated into computer programmes such as that found at www.gpfaqs.com or the Centre for Evidence-Based Medicine website, www.cebm.net) that enable health professionals to appraise for themselves the quality of the evidence. Assessing the quality of a randomised controlled trial may also be achieved using a Jadad score or other systems[14].

CRITICALLY APPRAISED TOPICS (CATS)

The result of performing a critical appraisal is a critically appraised topic or CAT (Fig. 1). These were introduced as people started to undertake critical appraisal[15]. They generally are a one page summary of all the key criteria of a trial. You can see that the key ingredients of a therapy trial are

Table II. Criteria for content, currency of information, and attribution and documentation of CATs.

Is the CAT valid?

1. Was the CAT focused by a well-built clinical question?
2. What was the explicit and sensible process used to identify and select the evidence?
3. Is it unlikely that relevant studies were missed?
4. Was the evidence appraised the best available to answer the question?
5. Were the appropriate validity criteria applied to the evidence appraised?
6. Are the dates clearly stated? Date of search, date of publication, date of expiry.

What is the CAT's message?

7. How strong is the message?
8. Is it expressed in terms likely to be helpful in clinical management?

Will the CAT's message help me in the care of my patients?

9. Can I apply the message in my patient setting, to my patients?
10. Were all clinically important outcomes, benefits, harms and costs discussed?

Criteria for authority of authors, disclosure of competing interests, and feedback mechanisms.

11. Is the academic or training level of the authors or commentators clearly stated?
12. Have the authors, site developers, and sponsors disclosed all competing interests?
13. Is there a mechanism to contact the authors?

[Acknowledgements to Victor Montori].

Table III. The ATTRACT process.

1. Clinician sends query to ATTRACT.
2. Researcher receives question and seeks clarification if necessary.
3. Researcher analyses terms – looks for synonyms, general classifications and MeSH terms – then devises a search strategy.
4. Researcher searches abstracts of evidence-based literature in accordance with the following hierarchy of sources:
   a. Cochrane Library
   b. Clinical Evidence (BMJ)
   c. TRIP database, including searching via clinical queries on PubMed
5. If a recent (less than 2 years old) systematic review is found that answers the query, the search will terminate here.
6. If no recent evidence is found from the core sources, then MEDLINE (via OVID or PubMed) and EMBASE are searched.
7. The research literature is appraised according to the standard EBM "Hierarchy of Evidence" and summarised.
8. A hardcopy of the summary and printouts of relevant research abstracts will be sent to the clinician who asked the question.
9. The summary will be uploaded on to the ATTRACT website as an answer to the question. It will be categorised according to subject area and, in cases where supporting evidence is weak, it will also be listed under poor research.

summed up in one sentence "double blinded randomised controlled trial (not clear if concealed) with intention to treat." The layout of all CATs varies slightly. For example, the journal Evidence-Based Medicine uses graphics to highlight the various sections of the trial design (Fig. 2).

Critical appraisal is not always straightforward. At first glance, it may appear quite simple as all that is required is to answer a list of questions from the paper and to enter the results into the online calculator to summarise the data. However, when we teach evidence-based practice, we find that it takes a considerable time to become confident in undertaking this[16,17]. The use of checklists is extremely useful and we teach medical students and residents how to appraise articles with these checklists. However, it does take some practice to get the data from the article correctly. For an inexperienced person, this usually means getting the CAT checked. In CAT banks, it is normal for the data to be checked by a second person.

As people started to undertake appraisal, an opportunity for sharing this work was realised[18]. Software has been written to facilitate the process. Most CAT software packages are actually web-based forms or word documents that the user then completes (c.f. the NHS Critical Appraisal Skills Programme or CASP, www.phru.nhs.uk/casp/casp.htm). To date, only one system automatically produces a CAT that is then published straight onto the web. This free pilot software is available at www.gpfaqs.com/.

There are EBM calculators that can perform the calculations required to produce, for example, a number needed to treat (NNT) and two good examples of these can be found at www.cebm.utoronto.ca/ or www.cebm.net/toolbox.asp. Not all CATs are equal and we must also have a way of assessing the reliability of the process. A suggested format is found in Table II.

There are three very different reasons for producing a CAT. The first and commonest is to appraise a recently-published article that is important. The second is to summarise an article that is being used to provide evidence as part of a guideline or textbook. The third is to answer an explicit clinical question.

The first is undertaken by publishers such as Evidence-Based Medicine and ACP Journal Club. These are gold standard CATs. The organisations providing these, employ professionals who appraise the article, another researcher validates the data, and an external commentator describes the relevance

Table IV . List and description of CAT banks.

CAT bank	URL	Brief description
EBM journal	http://ebm.bmjjournals.com/	Evidence-Based Medicine surveys a wide range of international medical journals applying strict criteria for the quality and validity of research. Practising clinicians assess the clinical relevance of the best studies. Articles are selected by physicians who ask whether the research will affect the way they practise or the way they think? The key details of these essential studies are presented in a succinct, informative abstract with an expert commentary on its clinical application. Published bi-monthly, Evidence-Based Medicine offers comprehensive coverage of primary care medicine. It includes a wide array of clinical disciplines, including family practice, internal medicine, paediatrics, obstetrics, gynaecology, psychiatry, and surgery.
ACP Journal Club	http://www.acpjc.org/	Similar to the EBM journal and appraised by the same team at McMaster University, it is run by the American College of Physicians. Relevance score and newsworthiness score are used to determine inclusion in the journal.
Evidence-Based On Call: EBOC	http://www.eboncall.org/	This project was carried out by a group working with the Centre for Evidence-Based Medicine in Oxford. It produced an emergency medicine textbook that links statements to CATs of the articles and is unique in this feature. It was finished in 2002 and now needs updating.
Inforetriever	http://www.infopoems.com/	POEMS (Patient Orientated Evidence that Matters) summarises an article and includes a commentary by a person in the field. Editors review more than 1,200 studies from over 100 journals monthly, presenting only the best as DailyPOEMs. One in 40 studies qualifies. The POEMs process applies specific criteria for validity and relevance to practise. The exact process of appraisal is less transparent.
Occupational Therapy CATS	http://www.otcats.com/topics/index.html	These are CATs completed by occupational therapists. Initially these were occupational therapists participating in a research project in 2002 and 2003.
Emergency Department, Manchester Royal Infirmary, UK	http://www.bestbets.org/	These are selected by practising emergency physicians from around the world. The search strategy is included and several articles are included in the answer. The search strategy is checked by an external reviewer. The CATs are very brief, usually just a paragraph, describing the study and results.
Royal Australian College of Physicians	http://www.racp.edu.au/imsanz/res_cat.htm	Very few CATs are publicly available on this site, 50 further CATs are available to RACP members.
Pediatric Critical Care Medicine CATs	http://pedsccm.wustl.edu/EBJournal_club.html	Peer-reviewed CATs of articles that are selected by clinicians. Trials selected for review are pertinent to practise.
University of North Carolina CATs	http://www.med.unc.edu/medicine/edursrc/!catlist.htm	Critical appraisal of articles selected by doctors. Most CATs are more than 5 years old.
University of Michigan Department of Pediatrics	http://www.med.umich.edu/pediatrics/ebm/cat.htm	Evidence-Based Pediatrics Web Site. This is a list of peer-reviewed CATs; the most recent is December 2004.
Critical Care Assembly Critically Appraised Topics	http://www.thoracic.org/criticalcare/cccat/library.asp	This is a list of peer-reviewed CATs expressed as paragraphs in a similar format to BestBETs.
National University Hospital CATs	http://www.nuh.com.sg/doctorhealth/ebm_catlibrary.htm	This library contains CATs prepared by physicians and residents of the National University Hospital. These CATs are not peer-reviewed.
Clinical Evidence	http://www.clinicalevidence.com/ceweb/conditions/index.jsp	The authors take a topic and provide the evidence for the statements. In a sense it is a clinical textbook but supported by evidence. The Clinical Evidence cites each article that it has appraised for validity. However these appraisals are not available, so if you want to check the details of the trial, you would have to go to the source. In Clinical Evidence, the date of the search of the cited article is provided.

of the article. The major benefit of these is that they include a commentary, usually by a clinician interested in the field. This puts the trial into context with other research as well as clinical practicalities. There are now journals for all the allied health professions, such as Evidence-Based Nursing, that undertake the same process.

The second is found in a few publications. Evidence-Based On Call (EBOC, www.eboncall.org) is one example where the evidence for each statement can be tracked backwards to the original CAT or CATs of articles that provide the evidence for that statement. EBOC provides both the summary as well as a CAT of the article. Clinical Evidence (www.clinicalevidence.com) is another example of this with the exception that the CAT is not visible to the reader. What is obvious from these publications is how quickly these types of research synthesis become out-of-date. For the EBOC data, the CATs all have a defined life that many have now exceeded.

The third is found in a few clinical answering services. This is by far the hardest process. It is the essence of evidence-based practice in that the health professional generates a clinical question from a real clinical situation[2]. The clinician then finds and appraises the evidence and finally applies this in practice. There has to be an explicit search of the literature describing where and how the search was performed. In addition, a reason for selection of the article(s) to be critically appraised must be given. The usual reason for selection of an article is based on the level of evidence (www.cebm.net/levels_of_evidence.asp) and the availability of the online full text of the article. Finally, the article or articles have to be appraised and an answer to the clinical question provided. This last reason was the initial purpose for creating CATs[19].

The ATTRACT (www.attract.wales.nhs.uk/) database is a question-answering system. It is run by professional health information specialists and has an extremely fast turnaround time for answering questions. It has a very explicit process by which questions are answered (Table III), but sadly this does not yet result in CATs.

Another answering system is FPIN (www.fpin.org). The concept of this site is that all health professionals are answering questions all the time and that if we share those answers, we would have a highly useful resource. Our actions are generally assessed in comparison to what others in our profession would have done in the same situation. With the exception of guidelines that often tend to be too generalised, it has been impossible to see what that consensus is unless we are unfortunate to end up in court. These sorts of question and answer databases allow us to see that consensus.

The FPIN database is run by family physicians from nine academic teaching sites in North America. It acts both as a highly academic answering system that provides teaching for the clinicians who provide the answers, as well as a database. It therefore has two functions in providing education on searching and appraisal, as well as providing clinical information. It is a slower process than the ATTRACT database, because it is being provided by health professionals working on a voluntary basis.

It takes too long for each clinician to do a search and appraise the answers to all clinical questions. As you can see from Table IV, people around the planet have already started systems to share CATs. This is the only effective approach to the information explosion in clinical medicine. We must share CATs on a global basis[18,20,21]. We are all asking the same questions and seeking the same answers. If we share that process, then we may be able to make the information needed by clinicians more widely available, as well as accessible at the point of clinical contact.

REFERENCES

1. Hook O. Scientific communications. History, electronic journals and impact factors. Scand J Rehabil Med 1999; 31:3-7.
2. Dawes M, Summerskill W, Glasziou P, et al. Sicily statement on evidence-based practice. BMC Med Educ 2005; 5: 1.
3. Hodgkin K. Diagnostic vocabulary for primary care. J Fam Pract 1979; 8: 129-44.
4. Gray JAM. Evidence-Based Healthcare. Churchill Livingstone, 1997: 17.
5. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001; 286:821-30.
6. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001; 323:42-6.
7. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282:1061-6.
8. Shea B, Moher D, Graham I, et al. A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. Eval Health Prof 2002; 25:116-29.
9. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999; 354:1896-900.
10. Makela M, Witt K. How to read a paper: critical appraisal of studies for application in healthcare. Singapore Med J 2005; 46:1-5.
11. Guyatt G, Rennie D, eds. Users' Guides to the Medical Literature: A Manual for Evidence-based Clinical Practice. Chicago: American Medical Association Press, 2002.
12. Sackett DL, Straus SE, Richardson WS, et al. Evidence-Based Medicine: How to Practice and Teach EBM. 2nd ed. Edinburgh: Churchill Livingstone, 2000.
13. Dawes M, Davis P, Gray A, et al. Evidence-Based Practice. A Primer for Health Care Professionals. 2nd ed. London: Elsevier, 2005.

14. Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials 1995; 16:62-73.
15. Sauve S, Lee HN, Meade MD, et al. The critically appraised topic: a practical approach to learning critical appraisal. Ann Roy Soc Phys Surg Canada 1995; 28:396-8.
16. Ramos KD, Schafer S, Tracz SM. Validation of the Fresno test of competence in evidence based medicine. BMJ 2003; 326:319-21.
17. Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. BMJ 2004; 329: 1017.
18. Dawes M, Godwin M. Global medical knowledge database is proposed. BMJ 2000; 320:1340.
19. Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the "evidence cart". JAMA 1998; 280:1336-8.
20. Dawes M. Creating a global medical knowledge database. The Family Doctor 2000; 16:16-7.
21. Dawes MG. On the need for evidence-based general and family practice. Evid Based Med 1996; 1:68-9.
22. Alkhenizan A. Diuretic based therapy reduced CV mortality in older patients with isolated systolic hypertension and diabetes. Evid Based Med 2005; 10:75.

SINGAPORE MEDICAL COUNCIL CATEGORY 3B CME PROGRAMME

Multiple Choice Questions (Code SMJ 200509A)

	True	False
Question 1. Indicate if the following statements on primary research are true or false:		
(a) Trials published in reputable journals need not be further appraised because they have undergone a rigorous editing process.	❑	❑
(b) Clinical outcomes in therapeutic trials are minimally influenced by trial methodology.	❑	❑
(c) Blinding (masking) and concealment of randomisation are important components of randomised controlled trials.	❑	❑
(d) Clinical trials published in primary research journals should be critically appraised by readers.	❑	❑
Question 2. The following are useful tools in producing a critically appraised topic (CAT):		
(a) Checklists for critical appraisal.	❑	❑
(b) EBM calculators (e.g. to perform calculations required to produce a Number Needed to Treat).	❑	❑
(c) CAT software packages.	❑	❑
(d) QUOROM statement checklist.	❑	❑
Question 3. The following are good reasons for producing a CAT:		
(a) To appraise an important article that has been recently published.	❑	❑
(b) To summarise an article that is being used to provide evidence for a set of clinical practice guidelines.	❑	❑
(c) To answer an explicit clinical question.	❑	❑
(d) To avoid having to perform a systematic review when there are multiple studies on the topic.	❑	❑
Question 4. We should ask the following questions when we read a CAT:		
(a) What was the explicit and sensible process used to identify and select the evidence?	❑	❑
(b) Will the CAT's message be applicable to my patient?	❑	❑
(c) Were all clinically important outcomes, benefits, harms and costs discussed?	❑	❑
(d) Have the authors disclosed all competing interests?	❑	❑
Question 5. The following are desirable features in a CAT:		
(a) Data in the CAT is validated by a second researcher.	❑	❑
(b) The CAT discusses the trial in its clinical context as well as clinical practicalities.	❑	❑
(c) It is shared widely so that others can benefit from it.	❑	❑
(d) It is house-trained and a good mouser.	❑	❑

Doctor's particulars:

Name in full: _____

MCR number: _____ Specialty: _____

Email address: _____

Submission instructions:

A. Using this answer form

1. Photocopy this answer form.
2. Indicate your responses by marking the "True" or "False" box ☑
3. Fill in your professional particulars.
4. Post the answer form to the SMJ at 2 College Road, Singapore 169850.

B. Electronic submission

1. Log on at the SMJ website: URL <http://www.sma.org.sg/cme/smj> and select the appropriate set of questions.
2. Select your answers and provide your name, email address and MCR number. Click on "Submit answers" to submit**.**

Deadline for submission: (September 2005 SMJ 3B CME programme): 12 noon, 25 October 2005

Results:

1. Answers will be published in the SMJ November 2005 issue.
2. The MCR numbers of successful candidates will be posted online at <http://www.sma.org.sg/cme/smj> by 20 November 2005.
3. All online submissions will receive an automatic email acknowledgment.
4. Passing mark is 60%. No mark will be deducted for incorrect answers.
5. The SMJ editorial office will submit the list of successful candidates to the Singapore Medical Council.