



Speed, accuracy, and confidence in Google, Ovid, PubMed, and UpToDate: results of a randomised trial

Robert H Thiele, Nathan C Poiró, David C Scalzo, et al.

Postgrad Med J 2010 86: 459-465
doi: 10.1136/pgmj.2010.098053

Updated information and services can be found at:
<http://pmj.bmj.com/content/86/1018/459.full.html>

These include:

References

This article cites 20 articles, 14 of which can be accessed free at:
<http://pmj.bmj.com/content/86/1018/459.full.html#ref-list-1>

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To order reprints of this article go to:
<http://pmj.bmj.com/cgi/reprintform>

To subscribe to *Postgraduate Medical Journal* go to:
<http://pmj.bmj.com/subscriptions>

Speed, accuracy, and confidence in Google, Ovid, PubMed, and UpToDate: results of a randomised trial

Robert H Thiele,¹ Nathan C Poiró,² David C Scalzo,¹ Edward C Nemergut³

► Additional material is published online only. To view these files please visit the journal online (<http://pmj.bmj.com>).

¹Department of Anaesthesiology, University of Virginia Health System, Charlottesville, Virginia, USA

²University of Virginia School of Medicine, University of Virginia Health System, Charlottesville, Virginia, USA

³Departments of Anesthesiology and Neurological Surgery, University of Virginia Health System, Charlottesville, Virginia, USA

Correspondence to

Robert H Thiele, Department of Anesthesiology, University of Virginia Health System, PO Box 800710, Charlottesville, VA 22908, USA; rht7w@virginia.edu

Received 25 January 2010
Accepted 10 May 2010

ABSTRACT

Background The explosion of biomedical information has led to an ‘information paradox’—the volume of biomedical information available has made it increasingly difficult to find relevant information when needed. It is thus increasingly critical for physicians to acquire a working knowledge of biomedical informatics.

Aim To evaluate four search tools commonly used to answer clinical questions, in terms of accuracy, speed, and user confidence.

Methods From December 2008 to June 2009, medical students, resident physicians, and attending physicians at the authors’ institution were asked to answer a set of four anaesthesia and/or critical care based clinical questions, within 5 min, using Google, Ovid, PubMed, or UpToDate (only one search tool per question). At the end of each search, participants rated their results on a four point confidence scale. One to 3 weeks after answering the initial four questions, users were randomised to one of the four search tools, and asked to answer eight questions, four of which were repeated. The primary outcome was defined as a correct answer with the highest level of confidence.

Results Google was the most popular search tool. Users of Google and UpToDate were more likely than users of PubMed to answer questions correctly. Subjects had the most confidence in UpToDate. Searches with Google and UpToDate were faster than searches with PubMed or Ovid.

Conclusion Non-Medline based search tools are not inferior to Medline based search tools for purposes of answering evidence based anaesthesia and critical care questions.

PURPOSE OF THE STUDY

The American Board of Medical Specialties (ABMS), which was founded by four medical speciality boards in 1933, now recognises 24 member boards¹ and issues more than 145 speciality and subspecialty certificates.² Medline, which contains references to over 18 million articles, added 1835 articles per day in 2007.³ From 1997 to 2010, the US National Institutes of Health budget grew from \$12.5 billion to over \$40 billion.⁴ The dual trends of superspecialisation in medicine and increased publicly available biomedical research have made it increasingly important for practising physicians to acquire and maintain a working knowledge of biomedical informatics.⁵ The purpose of this study was to evaluate four tools commonly used to answer clinical questions at our institution—Ovid, PubMed, Google, and UpToDate—in terms of accuracy, speed, and user confidence.

Ovid and PubMed were selected because both search the National Institutes of Health’s Medline

database (PubMed also searches the PreMedline database); however, they offer the user very different interfaces. Ovid requires multiple steps to initiate even the simplest query, whereas PubMed allows the user to initiate a query in one simple step (although more complex search strategies are available). Differences between the two can therefore be attributed to differences in the interface, rather than the searchable content.

Google was chosen because, while not specifically designed to search biomedical information, its highly sophisticated PageRank algorithm has made it the world’s most popular general search engine. Differences between internet based tools (eg, Google) and Medline based tools (eg, Ovid) can potentially lend insight into the relative importance of the body of work searched versus the search algorithms themselves.

UpToDate was chosen because, unlike the other three search tools, it is a proprietary depository of information generated by individual experts who are paid for their work. The culture of ‘evidence based medicine’ frowns on search strategies that do not involve looking for original answers themselves. From a practical (not educational) standpoint, this viewpoint is justifiable only if searches for original information are more successful than using proprietary collections of information such as UpToDate.

The utility of our study is based on four assumptions: first, that there are complex medical questions which have definitive, evidence based answers; second, that it is impossible for any individual to retain, at all times, the relevant biomedical information required to make informed medical decisions; third, that the average physician is unwilling to spend more than a few minutes searching for the information required to answer biomedical questions; and fourth, that there may be differences between tools used to navigate the biomedical literature.

STUDY DESIGN

Selection of eight clinical case scenarios

Two of the authors (RHT and ECN) have spent approximately 6 years reviewing the evidence behind many practice decisions commonly encountered by anaesthetists and intensivists.⁶ They chose eight articles which were felt to offer definitive answers to important clinical questions (for a list of these questions, and commentary on why they were selected, please see supplementary appendix) commonly faced by anaesthetists and intensivists.

Institutional review board approval, recruitment, and preparation

After acquiring approval from our institutional review board, we attempted to recruit 20 third and

fourth year medical students, 20 resident physicians, and 20 attending physicians at the University of Virginia. Subjects were told in advance which four search tools were included in the randomisation, and were asked to experiment with any they were unfamiliar with before the start of the trial.

Administration of clinical cases

Data collection

All searches were conducted on University Health System computers (which have access to virtually all electronically available primary literature) in the presence of a study coordinator, who recorded all search terms used and timed each search (for all questions, subjects were given a maximum of 5 min). At the end of each search, participants were asked by the study coordinator to rate their results on a 0–3 confidence scale (0 points were allotted for no answer): 1 point for an answer which they would not act on; 2 points for an answer which they would act on but only if they did not have additional time to search; and 3 points for an answer which they would act on confidently and which would require no additional searching. All data collection took place between December 2008 and June 2009.

Primary and secondary outcomes

The primary outcome was defined as determination of a correct answer with a confidence level of 3. Outcomes based on demographic variables, question difficulty, search engine preference, average time per question, and confidence in each engine (defined as the percentage of time in which a correct answer was assigned a confidence level of 3) were also explored, as were the effects of repeating a question. Additionally, the effect of the number of search terms on the various outcomes was analysed.

Part I: user selected search tools, randomised to four questions

All participants were initially randomised to a set of four (out of eight possible) clinical questions, which they were asked to answer in random order. Users were allowed to use any of the four pre-selected search tools (Ovid, PubMed, Google, UpToDate) to answer the initial four questions, and were allowed to use a different tool for each question, but were not allowed to use multiple tools in answering the same question.

Part II: randomised search tools, answered all eight questions

One to 3 weeks after answering the initial four questions, users were randomised to one of the four search tools, and asked to answer all eight questions (four of which they had previously answered) in random order. In some instances, subjects were randomised to the search tools they had originally chosen—this was allowed so that the effect of repeating a question could be delineated.

Data analysis

For data analysis, PubMed was arbitrarily chosen to be the 'reference standard'. Categorical frequency data were compared using χ^2 (with Yates' correction, as appropriate) unless otherwise noted. Multivariate comparisons to PubMed were made by an independent statistician who used logistic regression, multinomial logistic regression, and linear mixed models as appropriate.

RESULTS

Demographic differences

Fifteen medical students, 35 resident physicians, and four attending physicians completed the study (one attending physician withdrew after completing part I), answering a total

of 672 questions. All but two completed a demographic survey, the results of which are shown in table 1.

There were no statistically significant relationships between age, gender, or hours per week of computer use and primary outcome. Residents were more likely than medical students to achieve the primary outcome (32% vs 15%, $p < 0.001$, t test). There were not enough attending physicians to make any statistically significant inferences.

Question difficulty

There were significant differences in question difficulty, with questions 1 (inferior vena cava (IVC) filter and mortality difference, 71% answered correctly), 2 (appropriate tidal volume, 72% answered correctly), and 7 (transfusion threshold, 70% answered correctly) being easiest, and question 5 (bare metal stent and antiplatelet therapy before surgery, 27% answered correctly) being the most difficult ($p < 0.001$).

User preference

When given the choice, users chose Google 45% of the time, UpToDate 26% of the time, PubMed 25% of the time, and Ovid 4.4% of the time. This contrasts with the subjects' pre-test questionnaire, in which 33% claimed to use UpToDate most frequently, followed by Google (32%), and PubMed (13%).

Speed

Based on all 672 questions (parts I and II), there were significant differences in the speed of each search tool, with UpToDate, Google, PubMed, and Ovid taking 3.3, 3.8, 4.4, and 4.6 min, respectively ($p < 0.001$, one way analysis of variance (ANOVA)). On multivariate analysis (linear mixed model), both UpToDate and Google (average search times 3.3 and 3.8 min, respectively) were faster than PubMed (4.4 min, $p < 0.001$ and $p = 0.047$, respectively). Ovid (average search time 4.6 min) was significantly slower than PubMed ($p = 0.003$).

Table 1 Demographics and computer use of study participants

Average age	29.8 years
Gender	
Male	23 (43%)
Female	31 (57%)
Level of training	
Medical student	15 (28%)
Resident	35 (65%)
Attending	4 (7%)
Computer use	
Hours/week	17.0
Operating system preference	
Windows	41 (76%)
OSX	13 (24%)
Preferred internet browser	
Internet Explorer (Microsoft)	26 (48%)
Firefox (Mozilla)	22 (41%)
Other	6 (11%)
Preferred internet search engine	
Google	49 (91%)
Other	3 (9%)
Preferred biomedical search tool	
UpToDate	18 (33%)
Google	17 (32%)
PubMed	7 (13%)
Other	11 (22%)

Part I: user selected search tools, randomised to four questions Correct answers, confidence, and primary outcome

When users were allowed to choose their own search tool, users of both Google and UpToDate were more likely to find the correct answer ($p=0.0045$ and <0.001 , respectively) and achieve the primary outcome ($p=0.014$ and 0.030 , respectively). Differences between Ovid and PubMed did not achieve statistical significance for finding the correct answer ($p=0.39$) or achieving the primary outcome ($p=0.14$). There were no statistically significant differences in confidence between any of the search tools and PubMed ($p=0.14$, 0.24 , and 0.20 for Google, Ovid, and UpToDate, respectively) (figure 1).

Part II: randomised search tools

Effect of being randomised to a new search tool

In part II, subjects' ability to answer questions correctly was independent of whether or not they received the question in part I ($p=0.44$). Those who were assigned a new search tool to answer a previously asked question were more confident in their answer ($p=0.0015$) and more likely to achieve the primary outcome ($p<0.001$) than those who received a new question (figure 2).

Comparison when subjects were randomised to a new search tool

When subjects were randomised to a new search tool (ie, not the user's preference for that particular question in part I) for repeated questions in part II, users of Ovid were significantly less likely to find the correct answer ($p=0.015$) and achieve the primary outcome ($p=0.017$) as compared to users of PubMed. Otherwise, differences in subjects' ability to answer questions correctly or achieve the primary outcome did not achieve statistical significance on univariate analysis. There were no statistically significant differences in confidence between the four search tools (figure 3).

When using logistic regression to account for differences between individuals (those who answered correctly in part I were more likely to answer correctly in part II) and question difficulty, users of UpToDate were 2.76 times more likely (OR) than users of PubMed to achieve the primary outcome ($p=0.0015$, logistic regression). Differences between Google, Ovid, and PubMed did not achieve statistical significance.

Comparison for randomised search tool and new questions

Half of the questions in part II were new to the subject. When asked to answer new questions with no choice in what search

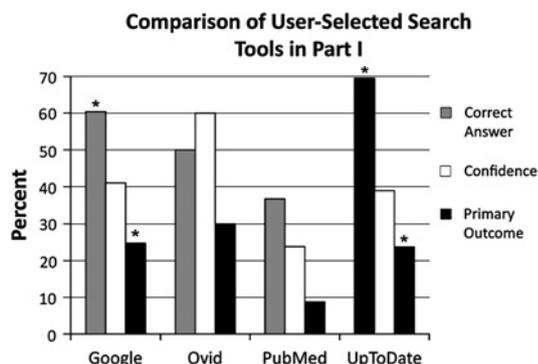


Figure 1 Chance (per cent) of determining the correct answer (black). Confidence, defined as the percentage of correct answers assigned a confidence level of three (white). Primary outcome, defined as a correct answer with a confidence level of three (grey). * $p<0.05$ compared to PubMed.

Effect of Being Randomized to a New Search Tool for Repeated Questions

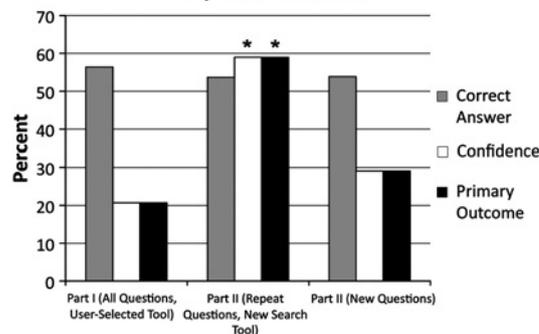


Figure 2 Users assigned to a search tool they did not choose in part I to answer repeated questions in part II were more confident in answering repeated questions than users randomised to new questions. * $p<0.05$ compared to part II (new questions).

tool to use (ie, randomised), subjects using Google and UpToDate were significantly more likely to find the correct answer than users assigned to PubMed ($p=0.05$ and 0.031 , respectively). Users randomised to UpToDate were significantly more confident in their answers ($p=0.007$). Users of UpToDate were also significantly more likely to achieve the primary outcome as compared to PubMed ($p<0.001$). Differences in achievement of the primary outcome between Google, Ovid, and PubMed did not achieve statistical significance (figure 4).

Pooled data from parts I and II

Primary outcome, univariate analysis, all questions

On univariate analysis, taking all questions into account, users of Google and UpToDate were more likely than users of PubMed to find the correct answer ($p=0.004$, <0.001 , respectively). Overall, users of UpToDate were significantly more confident in their answers ($p=0.005$) and more likely to achieve the primary outcome ($p<0.001$) than users of PubMed (figure 5).

In order to eliminate the possibility of bias due to repeating questions (mostly in the form of increased confidence), the data from parts I and II were re-analysed after removing all repeated questions. In this case, users of both Google and UpToDate were more likely than users of PubMed to find the correct answer ($p<0.001$ for both). Users of UpToDate were more confident than users of PubMed ($p<0.001$). Lastly, users of both Google and UpToDate were more likely than users of PubMed to

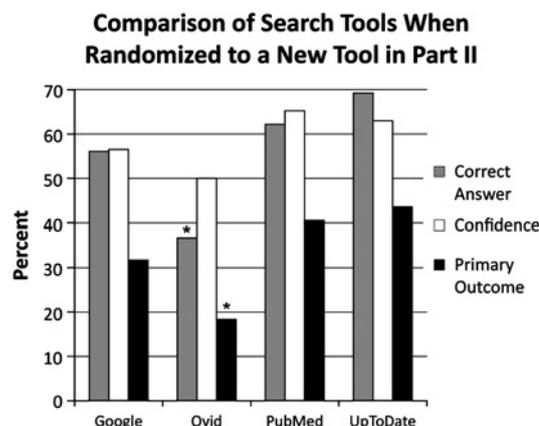


Figure 3 Users assigned to a new search tool (ie, not the search tool used in part I) for repeated questions in part II. * $p<0.05$ compared to PubMed.

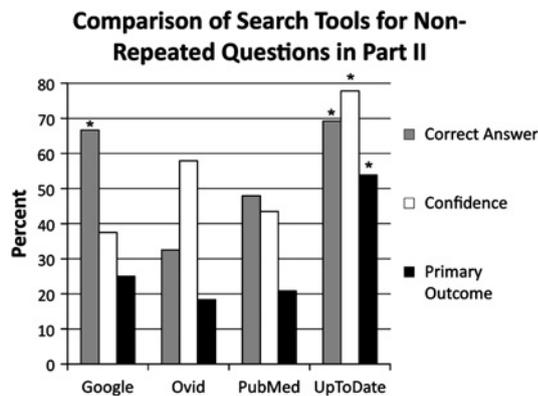


Figure 4 Users randomly assigned to search tool for previously unasked questions in part II. * $p < 0.05$ compared to PubMed.

achieve the primary outcome ($p = 0.03$, $p < 0.001$, respectively) (figure 6).

Confidence distribution

UpToDate users were significantly more likely than PubMed users to assign a confidence level of 3 to correct answers ($p < 0.001$) (figure 7). Ovid users were significantly more likely than PubMed users to not find an answer ($p < 0.001$), and both Google ($p = 0.025$) and UpToDate ($p = 0.011$) users were less likely than PubMed users to not find an answer (figure 8).

When adjusted for differences between individuals and question difficulty, UpToDate users were 3.29 times more likely (OR) than PubMed users to assign a confidence level 1 point higher (95% confidence limits 1.96 to 5.56, $p < 0.001$, multinomial logistic regression analysis). Users of Ovid were significantly less confident than users of PubMed (OR 0.3847, 0.1944 to 0.7613, $p = 0.0061$). There was a trend towards increased confidence among the Google users, although this did not achieve statistical significance.

Relationship between number of search terms and primary outcome

There was a trend towards an inverse relationship between primary outcome and the initial number of search terms used ($R^2 = 0.6872$, linear regression) (figure 9). There did not appear to be a relationship between the initial number of search terms and either the total time searching or the total number of searches ($R^2 = 0.03129$ and 0.33424 , respectively).

DISCUSSION

Our study showed that among a population of medical students, residents, and attending physicians asked to use Google, Ovid,

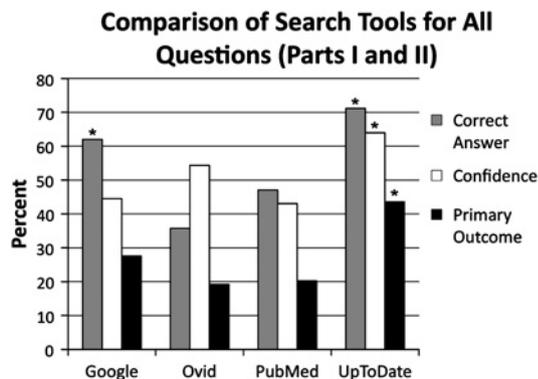


Figure 5 All data from parts I and II. * $p < 0.05$ compared to PubMed.

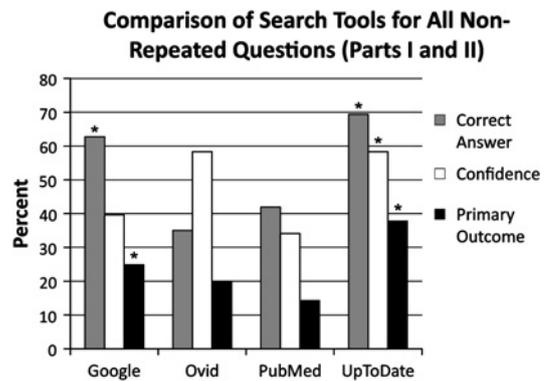


Figure 6 Data from part I and non-repeated questions in part II. * $p < 0.05$ compared to PubMed.

PubMed, and UpToDate to answer evidence based questions regarding the practice of anaesthesia or critical care medicine, those who choose Google and UpToDate are more likely to answer questions correctly than those who choose PubMed (our reference standard); that in addition to more commonly providing correct answers, searches with Google and UpToDate take less time than searches with PubMed; that users of UpToDate are significantly more confident in their answers than users of other search tools; that users who repeat a question are more confident in their answer despite no increased ability to answer the question; that when users are randomised to a new search tool, those who are randomised to Ovid are more likely to answer questions incorrectly than other search tools; and that users of Ovid are significantly more likely, and users of Google and UpToDate less likely, than users of PubMed to not find an answer.

The utility of these findings has been supported by several previous studies, most of which focus on the practice patterns of primary care physicians. Bates *et al* estimate that an ambulatory medical visit generates at least one clinical question that the clinician is unable to answer.⁷ In observing 103 primary care physicians (encompassing 2467 patient visits over 732 h), Ely *et al* found that 44% of encounters generated a clinically relevant question, and that of the 36% which were immediately pursued, physicians averaged less than 2 min of search time. Only two of 1101 questions led to a formal literature search.⁸

Seventy-two per cent of physicians report using the internet regularly for medical and professional updating,⁹ despite concerns regarding the quality of such information, which has been reviewed elsewhere.^{10–16} One early study showed relevance rates (number of relevant sites divided by the number of hits) ranging from 0.08 to 0.23.¹⁷ Physicians' willingness to use

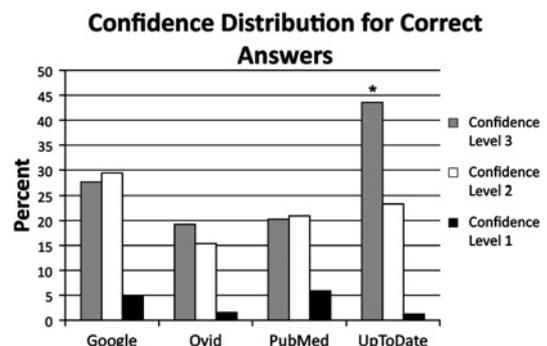


Figure 7 Distribution of confidence levels assigned to correct answers. * $p < 0.05$ compared to PubMed.

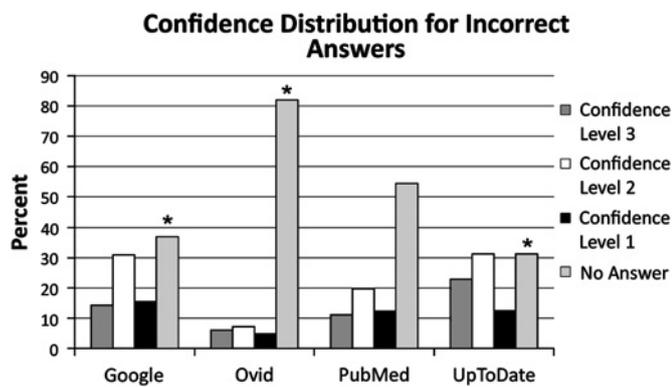


Figure 8 Distribution of confidence levels assigned to incorrect answers. * $p < 0.05$ compared to PubMed.

'questionable' sources of information (the internet) is likely a consequence of the commitment required to conduct a traditional literature search and the time constraints associated with modern medical practice. A study of questions asked by Missouri family physicians showed that formal Medline searches take an average of 27 min.¹⁸

While none of these studies addresses speciality or subspeciality physicians specifically, there is no reason to think that specialists (such as anaesthetists or intensivists) are not subject to similar practice pressures (increasing availability of information, limited time). Our results, which show that trainees and practising physicians only achieved the primary outcome 28% of the time, support a need for further understanding of the various search tools available to all physicians, not just primary care physicians.

Several authors have attempted to investigate biomedical search strategies formally, but the results are not necessarily applicable to the speciality practitioner. Ilic *et al* compared the ability of AltaVista, Excite, Google, Yahoo, and five medical search engines to find information about androgen deficiency, entering 18 keywords, phrases, or 'Boolean searches' into each tool, and examining the first 50 websites from each (a total of 9000 websites were examined). Results were assessed using the DISCERN quality assessment tool,¹⁹ with the 'medical' websites having a statistically insignificant tendency towards a higher

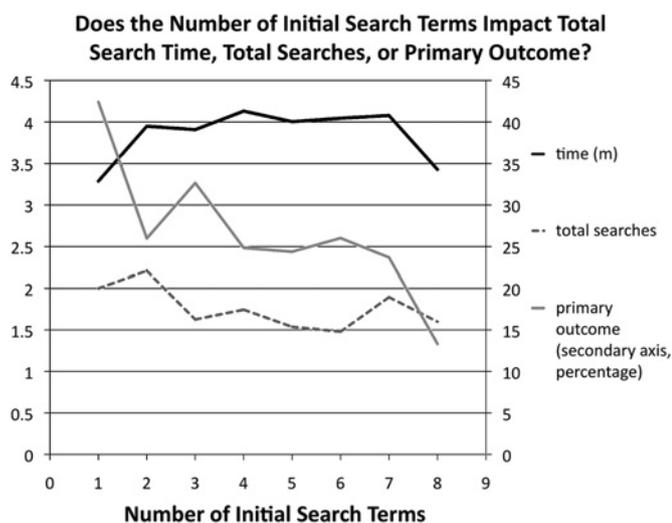


Figure 9 Relationship between number of initial search terms and total time per question (in minutes), total number of searches entered per question, and primary outcome (percentage).

DISCERN score. Of the non-medical websites, Google generated the highest percentage of relevant websites.²⁰ Ilic's study is limited by the fact that the authors (not users) created their own search terms.

Yu and Kaufman evaluated the ability of Google, MedQA, Onelook, and PubMed to answer definitional questions ('what is X?'). They presented four physicians (all of whom were biomedical informaticians) with 12 questions, and found that the subjective (user rated) 'quality of answer' was highest in Google. Furthermore, Google, MedQA, and Onelook were significantly faster than PubMed.²¹ The subjective nature of the study subjects' results (which were not verified) must be viewed with caution. Likewise, the four subjects' experience and training as bioinformaticians seriously threatens the broad applicability of the study's results.

Johnson *et al* created 10 questions 'designed to simulate... when a patient's clinical history includes a syndrome, medical device, or diagnostic test with which they are not familiar'. Medical students were given these 10 questions and asked to record the web resources used in the search as well as the number of links used. Students who started out with Google clicked on 0.44 fewer links before finding an answer than those who did not.²² Answers were not evaluated, time was not recorded, and searches were un-witnessed.

Tang and Ng attempted to show that search engines can facilitate diagnostic determination in difficult clinical cases. They reviewed 26 cases presented in the *New England Journal of Medicine*, and, before discovering the diagnosis, entered 3–5 search terms from each case into Google. After entering the search terms, the three diagnoses that best seemed to fit the symptoms and signs were selected. Using this methodology, Google found (but did not identify) the correct diagnosis in 58% of cases.²³

Our study was designed to be different from prior studies in several respects: first, we chose to study anaesthesia (speciality) providers, rather than generalists; second, we randomised our subjects; third, we scored our answers; fourth, we asked participants to report their confidence in each answer; fifth, all searches performed in this study were monitored by a study coordinator; and sixth, our questions were based on eight clinical case scenarios which are likely to arise in the daily practice of a speciality provider, and for which there was a highly defensible answer, supported by the biomedical literature (see supplementary appendix)—previous studies have proven the utility of search engines when the user is in need of obscure or rare pieces of information,²² situations in which an internet based search engine would be expected to outperform a standard text or journal (eg, 'what enzyme converts succinylCoA to d-aminolevulinic acid?'). We are not aware of any studies which combined these features.

The impetus for our study was the widespread belief among attending physicians at our institution that the use of non-Medline based search tools to answer biomedical questions at least partially invalidates the answer. This belief is not supported by any data of which we are aware. Attempts to compare online search tools to more traditional methods have been plagued by a combination of both methodological concerns (mentioned above) and the difficulty of studying a rapidly changing landscape.

Regarding the validity of Medline based search tools, some authors have begun to promote a biomedical 'knowledge hierarchy' with accompanying recommendations for optimising one's pursuit of knowledge. Interestingly (and surprisingly for some), primary sources and systemic reviews (two mainstays of

Medline based queries) make up the bottom two levels of the informational hierarchy. By contrast, evidence based texts such as UpToDate make up the fourth level of the hierarchy, second only to dedicated decision support systems in their validity.²⁴

In addition to the aforementioned results, our study produced several unexpected findings. We expected that subjects would be more successful at answering biomedical questions after repeating them. As shown in figure 2, subjects who answered repeated questions were more confident in their answers, but their ability to answer correctly was no different. This was not simply due to the fact that most subjects were randomised to a new search tool, as analysis of those randomised to the same search tool used in part I revealed only a slight improvement in their ability to find the answer of a repeated question (65% vs 70% for parts I and II, respectively), but a significant improvement in the primary outcome (21% vs 49% for parts I and II, respectively), primarily due to increased confidence in their answers. Thus, confidence appears to be more a function of how many times the question was asked, rather than what search engine was used, or whether or not the correct answer was found.

We were also surprised to find that subjects at our institution were significantly more confident in UpToDate than in PubMed (differences between Google, Ovid, and PubMed were not statistically significant). This is surprising given that Google (44.5%) was chosen more commonly than PubMed (25.1%), Ovid (4.41%), or UpToDate (26.0%). As (presumably) none of the 56 users were search-naïve, one would expect users to choose the search tools in which they had the most confidence.

Lastly, our study showed that when users are randomised (in part II) to a search engine that they had not initially chosen (in part I), differences between Google, PubMed, and UpToDate disappeared (Ovid still fared worse, in terms of both correct answers and primary outcomes). This is a key finding, because projected differences in a user's ability to find correct information and act on it are a reflection of both answer correctness and confidence in selected search tools. A unifying explanation for this result is difficult to find; the lack of difference between Google, PubMed, and UpToDate when users are randomly assigned to them for repeated questions, combined with the finding that answer correctness does not improve when users are allowed to select their own search tools, suggests that the effectiveness of a search tool is heavily dependent on the user's familiarity with it. That said, Google and UpToDate outperformed PubMed in subjects randomised to answer a novel question in part II. Thus, more data are needed to determine whether or not there is truly a difference between searches conducted by subjects who select their own search tool or are randomised to one of the four included in this study.

Limitations of our study, in addition to the suboptimal distribution of participants (the majority of whom were residents), include subjectivity in selecting the eight study questions (we intended to find eight critical care/perioperative questions which had a well supported answer), our inability to test all available search tools (eg, Wikipedia, NHS Evidence), and our inability to record accurately and reliably advanced features of the different search tools used (eg, Advanced Google search, Google Scholar vs Google, PubMed RCT designation, PubMed Clinical Queries, MeSH terms, etc). Furthermore, while we asked our participants to rank their answers based on anticipated behaviours, in the absence of an exceedingly complicated clinical study, there is no way to determine if our users would truly have acted as they reported. Our study relied on voluntary participation, and may have disproportionately drawn more technically gifted individuals. Lastly, we did not record or require

Main messages

- ▶ Subjects who are asked evidence based anaesthesia or critical care questions are more likely to answer them correctly when using Google or UpToDate as compared to PubMed.
- ▶ Among the four search tools evaluated in this study (Google, Ovid, PubMed, UpToDate), Google was the most popular but users were most confident in the results provided by UpToDate.
- ▶ Searches using Google and UpToDate are significantly faster than those conducted using Pubmed.
- ▶ There is no evidence that Medline based biomedical search tools are superior to alternatives such as Google or UpToDate.

Current research questions

- ▶ Are there any educational interventions which can improve search skills?
- ▶ Why, among the four tested search tools, do subjects choose Google most commonly, despite the fact that they are more confident in UpToDate.
- ▶ Are the results of this study applicable to other types of questions (eg, definitional).
- ▶ Which characteristics of Google, Ovid, PubMed, and UpToDate are beneficial, and which are detrimental?

primary literature sources in support of the answers (a decision we made because not everyone necessarily uses primary literature to make decisions—in retrospect we wish that we had recorded whether or not the end result of each search was a primary literature citation).

The major implication of our study is that while searching for original biomedical literature on Medline based search tools may offer educational value, for the purposes of finding correct information in a timely manner and with confidence, these tools (PubMed and Ovid) appear to be inferior to both Google and UpToDate.

CONCLUSION

Medical students and resident physicians are most likely to answer correctly anaesthesia and/or critical care based questions when using Google or UpToDate as compared to PubMed and Ovid. Furthermore, when using Google or UpToDate, they use significantly less time per question. Subjects were most confident in UpToDate, despite choosing Google most frequently. Subjects are not more likely to answer repeated questions correctly, although they are more confident in their answers when repeating questions. Subjects who are randomly assigned to a search tool are just as likely to answer questions correctly as those who are allowed to choose their search tool.

Acknowledgements The authors would like to thank the medical students, resident physicians, and attending physicians who donated their time in order to participate in this study

Funding Department of Anesthesiology, University of Virginia, PO Box 800710, Charlottesville, VA 22908-0710.

Competing interests None.

Ethics approval This study was conducted with the approval of the Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **ABA.** ABA Member Boards. http://www.abms.org/About_ABMS/who_we_are.aspx.
2. **ABA.** ABA Subspecialty Certificates. http://www.abms.org/About_ABMS/ABMS_History/Extended_History/Expansion.aspx.
3. **Medicine NLo.** Fact Sheet: Medline. 2008. <http://www.nlm.nih.gov/pubs/factsheets/medline.html> (accessed 10 Jul 2009).
4. **Steinbrook R.** The NIH stimulus—the recovery act and biomedical research. *N Engl J Med* 2009;**360**:1479–81.
5. **Gray JA.** Where's the chief knowledge officer? To manage the most precious resource of all. *BMJ* 1998;**317**:832.
6. **Open Anesthesia.** 2010 April 7. http://openanesthesia.org/index.php?title=Main_Page (accessed 2008).
7. **Bates DW,** Ebell M, Gottlieb E, *et al.* A proposal for electronic medical records in U.S. primary care. *J Am Med Inform Assoc* 2003;**10**:1–10.
8. **Ely JW,** Osheroff JA, Ebell MH, *et al.* Analysis of questions asked by family doctors regarding patient care. *BMJ* 1999;**319**:358–61.
9. **Podichetty VK,** Booher J, Whitfield M, *et al.* Assessment of internet use and effects among healthcare professionals: a cross sectional survey. *Postgrad Med J* 2006;**82**:274–9.
10. **Silberg WM,** Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: caveat lector et viewer—Let the reader and viewer beware. *JAMA* 1997;**277**:1244–5.
11. **Wyatt JC.** Commentary: measuring quality and impact of the World Wide Web. *BMJ* 1997;**314**:1879–81.
12. **Jadad AR,** Gagliardi A. Rating health information on the Internet: navigating to knowledge or to Babel? *JAMA* 1998;**279**:611–14.
13. **McClung HJ,** Murray RD, Heitlinger LA. The internet as a source for current patient information. *Pediatrics* 1998;**101**:E2.
14. **Griffiths KM,** Christensen H. Quality of web based information on treatment of depression: cross sectional survey. *BMJ* 2000;**321**:1511–15.
15. **Purcell GP,** Wilson P, Delamothe T. The quality of health information on the internet. *BMJ* 2002;**324**:557–8.
16. **Childs S.** Judging the quality of internet-based health information. *Performance Measurement and Metrics* 2005;**6**:80–96.
17. **Sacchetti P,** Zvara P, Plante MK. The Internet and patient education—resources and their reliability: focus on a select urologic topic. *Urology* 1999;**53**:1117–20.
18. **Chambliss ML,** Conley J. Answering clinical questions. *J Fam Pract* 1996;**43**:140–4.
19. **Charnock D,** Shepperd S, Needham G, *et al.* An instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 1999;**53**:105–11.
20. **Ilic D,** Bessell TL, Silagy CA, *et al.* Specialized medical search-engines are no better than general search-engines in sourcing consumer information about androgen deficiency. *Hum Reprod* 2003;**18**:557–61.
21. **Yu H,** Kaufman D. A cognitive evaluation of four online search engines for answering definitional questions posed by physicians. *Pac Symp Biocomput* 2007;**12**:328–39.
22. **Johnson PT,** Chen JK, Eng J, *et al.* A comparison of world wide web resources for identifying medical information. *Acad Radiol* 2008;**15**:1165–72.
23. **Tang H,** Ng JH. Googling for a diagnosis—use of Google as a diagnostic aid: internet based study. *BMJ* 2006;**333**:1143–5.
24. **Haynes B.** Of studies, syntheses, synopses, summaries, and systems: the “5S” evolution of information services for evidence-based healthcare decisions. *Evid Based Nurs* 2007;**10**:6–7.