Googling for a diagnosis—use of Google as a diagnostic aid: internet based study

Hangwi Tang, Jennifer Hwee Kwoon Ng

Abstract

Objective To determine how often searching with Google (the most popular search engine on the world wide web) leads doctors to the correct diagnosis.
Design Internet based study using Google to search for diagnoses; researchers were blind to the correct diagnoses.
Setting One year's (2005) diagnostic cases published in the case records of the *New England Journal of Medicine*.
Cases 26 cases from the *New England Journal of Medicine*; management cases were excluded.
Main outcome measure Percentage of correct diagnoses from Google searches (compared with the diagnoses as published in the *New England Journal of Medicine*).
Results Google searches revealed the correct diagnosis in 15 (58%, 95% confidence interval 38% to 77%) cases.
Conclusion As internet access becomes more readily available in outpatient clinics and hospital wards, the web is rapidly becoming an important clinical tool for doctors. The use of web based searching may help doctors to diagnose difficult cases.

Introduction

Doctors adept at using the internet use Google to help them diagnose difficult cases. As described in the *New England Journal of Medicine*,[1] a doctor astonished her colleagues (including an eminent professor) by correctly diagnosing IPEX (immunodeficiency, polyendocrinopathy, enteropathy, X linked) syndrome. She admitted that the diagnosis "popped right out" after she entered the salient features into Google.

It seems that patients use Google to diagnose their own medical disorders too. After evaluating a 16 year old water polo player who presented with acute subclavian vein thrombosis, one of us (HT) started to explain that the cause of the thrombosis was uncertain when the patient's father blurted out, "But of course he has Paget-von Schrötter syndrome." Having previously googled the symptoms, he gave us a mini-tutorial on the pathophysiology (hypertrophy of the neck muscles leading to dynamic compression of the axillary vein at the thoracic inlet—leading to thrombosis) and the correct treatment of the syndrome.[2] This experience led us to ask: "How good is Google in helping doctors to reach the correct diagnosis?"

Method

We selected a convenient sample of one year's (2005) diagnostic cases presented in the case records of the *New England Journal of Medicine*. We excluded management cases. After discussion, we selected three to five search terms from each case record and entered them on a data sheet. We then did a Google search for each case while blind to the correct diagnoses (that is, before reading the differential diagnosis and conclusion of each case record). We selected and recorded the three most prominent diagnoses that seemed to fit the symptoms and signs. We then compared the results with the correct diagnoses as published in the case records.

Results

We identified 26 cases from the case records (table 1). Google searches found the correct diagnosis in 15 (58%, 95% confidence interval 38% to 77%) cases. In some cases (for example, case record 9), Google gave the correct diagnosis (extrinsic allergic alveolitis) but we felt that it was not specific enough to be considered correct (extrinsic allergic alveolitis caused by *Mycobacterium avium*, also known as "hot tub lung").

Discussion

Clinical decision support programs have been reported to be valuable aids in diagnosing difficult cases.[3] Hoffer reported using a clinical decision support program to make the diagnosis of Addison's disease expeditiously when it was missed by many expert clinicians.[4 5] We think that Google is likely to be a useful aid in diagnosis too. It has the advantage of being easier to use and is freely available on the internet.

A few limitations of this study should be mentioned. Arguably, everything could be found on

P+ *An extra table is on bmj.com*

This article was posted on bmj.com on 10 November 2006: http://bmj.com/cgi/doi/10.1136/bmj.39003.640567.AE

Editorial by Gardner

Department of Respiratory and Sleep Medicine, Princess Alexandra Hospital, Brisbane, Q4102, Australia
Hangwi Tang *respiratory and sleep physician*
Department of Rheumatology, Princess Alexandra Hospital
Jennifer Hwee Kwoon Ng *consultant rheumatologist*
Correspondence to: H Tang hangwitang@yahoo.com

BMJ 2006;333:1143–5

the web if only one knew the correct search terms. In this case, we chose a combination of search terms that we felt would be unique (see extra table on bmj.com). We chose between three to five search terms for each case, depending on symptoms and signs that we felt would not return a non-specific result. We selected "statistically improbable phrases" whenever possible,[6] such as "cardiac arrest sleep" in case record 37. We generally selected likely diagnoses from the first three pages (maximum five pages) of the search result, containing 30 documents, to see if the condition would fit the case record. As Google does not "suggest" a diagnosis, we selected the diagnosis that we felt would fit best with the case record. When none of the diagnoses found with Google fitted the case record well, we chose up to three most likely diagnoses. If one of the diagnoses was correct, we regarded the search as successful.

We suspect that using Google to search for a diagnosis is likely to be more effective for conditions with unique symptoms and signs that can easily be used as search terms, such as the one described by Greenwald.[1] Searches are less likely to be successful in complex diseases with non-specific symptoms (case records 10 and 14) or common diseases with rare presentations (case record 18).

The efficiency of the search and the usefulness of the retrieved information also depend on the searchers' knowledge base. In this case, although we were blinded to the correct diagnosis, one author was a respiratory and sleep trainee and the other a rheumatologist; sometimes the diagnoses were evident to us, and this could have affected our choice of search terms. When choosing the "correct" diagnoses from a list of possible choices returned by Google, we tried to avoid using specialist knowledge but chose diagnoses that were ranked most prominently and seemed to fit the case record. Therefore, for case record 9, where we made the correct diagnosis of "hot tub lung," searching with Google did not give enough prominence to hot tub lung for it to be considered the correct answer.

Google diagnoses and actual diagnoses for 26 case reports

Case record	Google diagnosis	Final diagnosis	Google diagnosis correct?
5	Infective endocarditis	Infective endocarditis	Yes
6	Gastrointestinal bleed	Linitis plastica with bowel obstruction	No
7	Cushing's syndrome	Cushing's syndrome secondary to adrenal adenoma	Yes
8	Eosinophilic granuloma, osteoid osteoma	Osteoid osteoma	Yes
9	Extrinsic allergic alveolitis, tuberculosis, BOOP	Hot tub lung secondary to *Mycobacterium avium*	No
10	Amyotrophy	Ehrlichiosis	No
11	Tuberculosis, lymphoma	Lymphoma	Yes
12	Neurofibromatosis type 1	Neurofibromatosis type 1	Yes
14	Uveitis	Vasculitis	No
15	Amyloid	Amyloid light chain	Yes
16	Hyperaldosteronism	Phaeochromocytoma	No
17	Acute chest syndrome	Acute chest syndrome	Yes
18	Tuberous sclerosis	Endometriosis	No
19	Aspergillus	Aspiration pneumonia, brain abscess	No
22	Graft versus host disease	West Nile fever	No
25	Cirrhosis	Pylephlebitis	No
26	Hypertrophic obstructive cardiomyopathy	Hypertrophic obstructive cardiomyopathy	Yes
27	Spongiform encephalopathy (Creutzfeldt-Jakob disease)	Creutzfeldt-Jakob disease	Yes
28	Churg-Strauss syndrome	Churg-Strauss syndrome	Yes
29	Polymyositis or dermatomyositis	Dermatomyositis secondary to non-Hodgkin's lymphoma	Yes
30	Cat scratch disease	Cat scratch disease	Yes
31	Henoch-Schonlein purpura	Cryoglobulinaemia	No
33	First hit=juvenile polyposis plus HTT, which links to MADH4 mutation	MADH4 mutation (HTT plus juvenile polyposis)	Yes
34	Toxic epidermal necrolysis syndrome	Toxic epidermal necrolysis syndrome	Yes
36	Encephalitis	MELAS	No
37	Long QT syndrome, Brugada syndrome	Brugada syndrome	Yes

BOOP=bronchiolitis obliterans organising pneumonia; HTT=hereditary haemorrhagic telangiectasia; MELAS=myoclonus epilepsy lactic acidosis stroke-like syndrome.

What is already known on this topic

Doctors and patients are increasingly using the internet to search for health related information

Google is the most popular search engine on the world wide web

What this study adds

Searching with Google may help doctors to formulate a differential diagnosis in difficult diagnostic cases

Patients doing a Google search may find the search less efficient and be less likely to reach the correct diagnosis. We believe that Google searches by a "human expert" (a doctor) have a better yield, as Google is exceedingly good at finding documents with co-occurrence of the signs/symptoms used as search terms and human experts are efficient in selecting relevant documents. Furthermore, doctors in training would find the Google searches educational and useful in formulating a differential diagnoses.

The role of diagnostician remains one of the most challenging and fulfilling roles of a physician. Physicians have been estimated to carry two million facts in their heads to fulfil this role.[7] With medical knowledge expanding rapidly, even this may not be enough. Search engines allow quick access to an ever increasing knowledge base.[8] Google gives users ready access to more than three billion articles on the web[9] and has far exceeded PubMed as the search engine of choice for retrieving medical articles.[10] Google has been so popular that the word has entered the English lexicon as a verb.[11] Google Scholar, currently in beta form (www.scholar.google.com), is likely to be even more useful as it searches only peer reviewed articles.

Conclusions

Doctors and patients are increasing proficient with the internet and frequently use Google to search for medical information. Twenty five million people in the United Kingdom were estimated to have web access in 2001, and searching for health information was one of the most common uses of the web.[12] Computers connected to the internet are now ubiquitous in outpatient clinics and hospital wards. Useful information on even the rarest medical syndromes can now be found and digested within a matter of minutes. Our study suggests that in difficult diagnostic cases, it is often useful to "google for a diagnosis." Web based search engines such as Google are becoming the latest

tools in clinical medicine, and doctors in training need to become proficient in their use.

Contributors: HT had the idea and designed the study. JHKN helped in the study design. Both authors did the search and analysis and wrote the paper. HT is the guarantor.

Funding: None.

Competing interests: None declared.

Ethical approval: Not sought. The subjects were published cases in the *New England Journal of Medicine* with no patient identifiers.

1 Greenwald R. And a diagnostic test was performed. *N Engl J Med* 2005;359:2089-90.
2 Chaudhry MA, Hajarnavis J. Paget-von Schrötter syndrome: primary subclavian-axillary vein thrombosis in sport activities. *Clin J Sport Med* 2003;13:269-71.
3 Scott C. Diagnosing childhood conditions: have you considered...? *Medical Protection Society Casebook* 2005;13:22-5.
4 Hoffer EP. Clinical problem solving: identifying Addison's disease. *N Engl J Med* 1996;334:1403-5.
5 Keljo DJ, Squires RH. Clinical problem-solving: just in time. *N Engl J Med* 1996;334:46-8.
6 Wikipedia. Free encyclopedia. http://en.wikipedia.org/wiki/Main_Page (accessed 27 Jun 2006).
7 Pauker SG, Gorry GA, Kassirer JP, Schwartz WB. Towards the simulation of clinical cognition taking a present illness by computer. *Am J Med* 1976;60:981-96.
8 Giustini D. How Google is changing medicine. *BMJ* 2005;331:1487-8.
9 Vise D, Malseed M. *The Google story*. New York: Delacorte Press, 2005.
10 Steinbrook R. Searching for the right search—reaching the medical literature. *N Engl J Med* 2006;354:4-7.
11 *Oxford advanced learner's dictionary of current English*. 7th ed. New York: Oxford University Press, 2005.
12 Powell J, Clarke A. The www of the world wide web: who, what, and why? *J Med Internet Res* 2002;4(1):e4.

(Accepted 22 September 2006)

doi 10.1136/bmj.39003.640567.AE

Lifetime cost effectiveness of simvastatin in a range of risk groups and age groups derived from randomised trial of 20 536 people

Heart Protection Study Collaborative Group

Abstract

Objectives To evaluate the cost effectiveness of 40 mg simvastatin daily continued for life in people of different ages with differing risks of vascular disease.
Design A model developed from a randomised trial was used to estimate lifetime risks of vascular events and costs of treatment and hospital admissions in the United Kingdom.
Setting 69 hospitals in the UK.
Participants 20 536 men and women (aged 40-80) with coronary disease, other occlusive arterial disease, or diabetes.
Interventions 40 mg simvastatin daily versus placebo for an average of 5 years.
Main outcome measures Cost effectiveness of 40 mg simvastatin daily expressed as additional cost per life year gained. Major vascular event defined as non-fatal myocardial infarction or death from coronary disease, any stroke, or revascularisation procedure. Results were extrapolated to younger and older age groups at lower risk of vascular disease than were studied directly, as well as to lifetime treatment.
Results At the April 2005 UK price of £4.87 (€7; $9) per 28 day pack of generic 40 mg simvastatin, lifetime treatment was cost saving in most age groups and vascular disease risk groups studied directly. Gains in life expectancy and cost savings decreased with increasing age and with decreasing risk of vascular disease. People aged 40-49 with 5 year risks of major vascular events of 42% and 12% at start of treatment gained 2.49 and 1.67 life years, respectively. Treatment with statins remained cost saving or cost less than £2500 per life year gained in people as young as 35 years or as old as 85 with 5 year risks of a major vascular event as low as 5% at the start of treatment.
Conclusions Treatment with statins is cost effective in a wider population than is routinely treated at present.

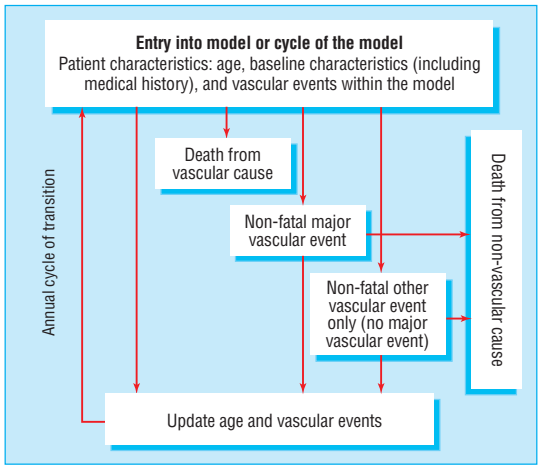

Fig 1 Schematic of the state transition model

Introduction

Large randomised trials have shown that lowering blood concentrations of low density lipoprotein cholesterol with statins greatly reduces rates of heart attacks, strokes, and revascularisation procedures in a wide range of people at high risk, largely irrespective of their cholesterol concentrations and other characteristics at presentation.[1] The heart protection study has shown that, especially when cheaper generic versions are used, 40 mg simvastatin daily is cost effective for a wider range of people with vascular disease or diabetes than previously thought.[2]

P+ *A table, a technical appendix, and details of collaborators, participating hospitals, and committees are on bmj.com*

ELPS *This is the abridged version of an article that was posted on bmj.com on 10 November 2006: http://bmj.com/cgi/doi/10.1136/bmj.38993.731725.BE*

Correspondence to:
hps@ctsu.ox.ac.uk

BMJ 2006;333:1145–8