A simple heuristic for Internet-based evidence search in primary care: a randomized controlled trial

Andreas Eberbach\(^1\)
Annette Becker\(^1\)
Justine Rochon\(^2\)
Holger Finkemeler\(^1\)
Achim Wagner\(^3\)
Norbert Donner-Banzhoff\(^4\)

\(^1\)Department of Family and Community Medicine, Philipp University of Marburg, Marburg, Germany; \(^2\)Institute of Medical Biometry and Informatics, University of Heidelberg, Heidelberg, Germany; \(^3\)Department of Sport Medicine, Justus-Liebig-University of Giessen, Giessen, Germany

Background: General practitioners (GPs) are confronted with a wide variety of clinical questions, many of which remain unanswered.

Methods: In order to assist GPs in finding quick, evidence-based answers, we developed a learning program (LP) with a short interactive workshop based on a simple three-step-heuristic to improve their search and appraisal competence (SAC). We evaluated the LP effectiveness with a randomized controlled trial (RCT). Participants (intervention group [IG] n=20; control group [CG] n=31) rated acceptance and satisfaction and also answered 39 knowledge questions to assess their SAC. We controlled for previous knowledge in content areas covered by the test.

Results: Main outcome – SAC: within both groups, the pre–post test shows significant (P<0.00) improvements in correctness (IG 15% vs CG 11%) and confidence (32% vs 26%) to find evidence-based answers. However, the SAC difference was not significant in the RCT.

Other measures: Most workshop participants rated “learning atmosphere” (90%), “skills acquired” (90%), and “relevancy to my practice” (86%) as good or very good. The LP-recommendations were implemented by 67% of the IG, whereas 15% of the CG already recommended LP recommendations spontaneously (odds ratio 9.6, P<0.00). After literature search, the IG showed a (not significantly) higher satisfaction regarding “time spent” (IG 80% vs CG 65%), “quality of information” (65% vs 54%), and “amount of information” (53% vs 47%).

Conclusion: Long-standing established GPs have a good SAC. Despite high acceptance, strong learning effects, positive search experience, and significant increase of SAC in the pre–post test, the RCT of our LP showed no significant difference in SAC between IG and CG. However, we suggest that our simple decision heuristic merits further investigation.

Keywords: decision making, medical informatics, evidence-based practice, continuing medical education

Introduction

Unanswered questions

In primary care, patients present a wide spectrum of clinical problems to their general practitioners (GPs). The questions that are brought up in these encounters touch upon virtually every clinical topic.\(^1\)-\(^3\) According to a recent systematic review, ten patient visits left GPs with approximately six unanswered questions on average. Only ~50% of GPs attempted to find an answer.\(^4\) The main reasons given by GPs for not pursuing their open questions further were lack of time and doubt that relevant information was available at all.
Available information

There is a discrepancy between the situation described earlier and the large amount of medical information available. Structured databases and other sources are accessible via the Internet. Online evidence has the potential to provide quick answers to physicians’ questions.\textsuperscript{5}

Clinicians’ use of electronic resources has been reported to be increasing.\textsuperscript{6} Examinations of resources used in primary care have revealed that, after textbooks and discussions with other physicians, Internet use represents the third most frequently used resource.\textsuperscript{4,6–10}

Answering questions

Evidence-based medicine (EbM) has provided a framework for accessing, appraising, and using the medical literature relevant to individual patients. One of its pioneers, David Sackett, coined the paradigm of “Five Steps”: question formulation, literature search, critical appraisal, implementation of results, and evaluation.\textsuperscript{11} However, time constraints and lack of skills prevent most physicians from following this model.\textsuperscript{1,10,12–16}

As an alternative to the classical appraising mode, the McMaster Group already suggested the searching mode, which primarily uses databases of processed information for quick answers to questions arising in everyday care situations. Even physicians familiar with EbM appraisal methods would resort to less sophisticated strategies for their everyday clinical questions. Slawson and Shaughnessy have formulated a need for simple heuristics not just to appraise but to manage medical information.\textsuperscript{17,18}

Methods

Learning program (LP)

We developed an LP that is based on a simple three-step heuristic (Figure 1). The aim of the program is to help GPs use Internet sources for their clinical questions. This can be taught in interactive seminars. All information is also available in the form of a booklet for individual study. A CD-ROM with relevant URLs and background material helps GPs put into practice what they have learned. In order to improve the effectiveness of the program, we used multimedia (workshop [WS], booklet, CD, Internet) and multiple exposures (lectures, exercises, homework).\textsuperscript{19}

Simple heuristic

The content of the LP (Figure 2) was structured around the three-step heuristic,\textsuperscript{20} shown in Figure 1. Its application ensures that GPs search and appraise sources in a methodological way and also helps them decide when to end their search. We regard this as especially important because GPs often feel overwhelmed by the amount of available information.\textsuperscript{21–23} This set of three questions is not meant to be followed strictly in every literature search. It is also meant to encourage GPs to reflect on their own search and appraisal habits.

Example 1: a commercial manufacturer mentions in the drug-sheet of a lipid-lowering drug that no long-term studies with patient-relevant outcomes have been conducted so far.

![Figure 1 The three-step heuristic of decision making.](https://www.dovepress.com/)

**Table of contents**

- Formulating "researchable" questions
- The way of knowledge
- All sources are biased: let's use them!
- Intermission: Searching the Internet
- Searching and finding
- Printing, collecting, saving
- Finding and using sources
- Our concept for evaluating selected online sources
- Selected online sources: Drugs
- Selected online sources: Clinical practice guidelines
- Selected online sources: Databases and professional journals
- On professional journals on the Internet
- Analyzing and evaluating retrieved information
- Effectiveness trials
- Study designs
- Validity criteria
- Therapy assessment (ARR, RRR, NNT)
- Statistical significance, clinical relevance
- Summing up: three steps
- Information management for general practitioners
- Old and new
- On "rent-yappers", sponsored journals
- ... and other forms of landscape management
- Taking questions seriously
- I've developed an appetite – and now?

![Figure 2 Workshop table of contents. Abbreviations: ARR, absolute risk reduction; RRR, relative risk reduction; NNT, number needed to treat.](https://www.dovepress.com/)
This information is not in line with the interest of the manufacturer (source); it should therefore be accepted without further search. The claim of a manufacturer that a herbal preparation alleviates tumor pain, however, would not lead to immediate acceptance but should motivate further research.

Example 2: if for a herbal preparation, no formal evaluation, such as a controlled study, can be found, its use should be avoided. Only if the claim is in line with the interest of the source and trials evaluating a particular treatment exist, a more detailed appraisal of study design features is required. Here the most valid and relevant criteria should be checked first, such as:

- is this a randomized controlled trial (RCT)?
- is the main outcome clinically relevant?
- can the choice of the control treatment be justified?
- which patients have been recruited into the study (external generalizability)?

Participants

GPs located in five different cities in Germany were invited to take part in the main study. Participants had to be motivated to attend two WS sessions and contribute to the evaluation.

Stratification

GPs were asked whether they had performed an Internet search for a clinical question during the previous year. The answer (yes/no) was used for stratification.

Randomization

The computer-assisted randomization was undertaken by the central clinical trials unit at the University of Marburg. Within the two stratification-groups, permuted blocks of variable length were formed, including one to four participants. According to the order of registration, the GPs were assigned to either the intervention group (IG) or the control group (CG). In order to maintain the allocation concealment, the participants were not given details about the study design or the allocation process. The control arm was a waiting group of GPs who were offered the WS after study completion and evaluation of the main outcome.

Intervention

GPs attended two interactive small group WS in a computer classroom. Figure 2 shows the contents of the WS. Sessions provided information on databases that are useful for GPs. Their scope, advantages, and disadvantages were discussed. There was an emphasis on processed evidence-based information as opposed to original papers, such as drug bulletins, clinical evidence (British Medical Journal) or national and international guidelines. Practical search strategies, bias of sources, and simple criteria to check the validity of information were covered. There was ample opportunity for hands-on exercises and discussion of participants’ experience with the sources provided. WS lasted at least 2 hours each with a 4-week interval in-between. More specific background information can be found in the “Supplementary materials” section (English) and on the Internet (German).

Study outcomes

The primary outcome of the study was GPs’ search and appraisal competence (SAC). The secondary outcome was to determine the acceptance and satisfaction with the WS.

WS acceptance

We asked GPs to evaluate both WS using a semi-standardized questionnaire each time directly after the event. Fifteen questions in a 6-point Likert scale covered organization, educational process, and relevance of content (1= very good, 6= deficient). Finally, they estimated their own commitment for change regarding content covered by the WS in a standardized way.

Search satisfaction

GPs were asked to rate the relevance of content and their satisfaction using the strategies learnt. Acceptance and satisfaction were evaluated not only by IG participants but also by control subjects after their participation in the WS.

SAC

We evaluated GPs’ SAC by additionally (after pre–post test) correct answered clinical questions.

Clinical questions

Thirty-nine clinical questions from seven clinical domains were developed by the study team. Questions concerned prevention, therapy, diagnosis, and prognosis. Most questions were in multiple-choice format, but GPs were also asked to elaborate on some open questions. The appropriateness of answers was checked by experts in EbM and clinical areas covered (see the “Acknowledgments” section).

Pre–post test

All participants had to answer each question twice: first, spontaneously in order to establish their baseline knowledge and again after a thorough search. We could thus control for previous knowledge.
If participants improved, ie, gave an incorrect answer at baseline but a correct answer after thorough search, the answer was scored “+1”. In the case of a correct answer at baseline but a wrong answer after the search, the score was “–1”; no change resulted in “0”. We arrived at an overall SAC score by adding up the single question scores for each participating GP. The result can theoretically vary between −39 (−100%) and +39 (+100%) additional correct answers. Besides the objective professional correctness, we checked the subjective confidence (0%–100%).

**Pilot study**

Before starting the study, we checked the comprehensibility and the level of difficulty of the 39 clinical questions with a group of 12 GPs. Subsequently, we held the workshop (WS) with another 14 GPs to improve methods and content including learning material. These GPs did not take part in the main study.

**Statistics**

The data analyses were performed with SPSS software (SPSS Inc., Chicago, IL, USA). For baseline comparisons, we report medians, interquartile ranges, and results of the Mann–Whitney U test (P-values) for continuous variables and percentages, and results of asymptotic χ² tests for categorical variables. Differences in SAC improvement (before vs after; between study arms) were evaluated by the Mann–Whitney U test (P-value). We used the Hodges–Lehmann estimator to obtain the 95% confidence interval for the difference between medians.26

**Results**

**Study participants**

Recruitment, randomization, and flow of participants are shown in Figure 3. Twelve percent (n=110 of 905) of the GPs originally approached gave written consent to participate and were randomized thereafter. Seventy-six percent (n=84 of 110) of subscribers attended at least one WS. Finally, complete data were available from 46% (n=51 of 110) of the original group. Consequently, the dropout rate was 64% due to delayed information search (n=30 “did not search”), absence (n=10 “did not attend”), and incomplete WS attendance (n=15 “only one WS”) or incomplete data (n=4).

![Figure 3](https://www.dovepress.com/)

**Figure 3** Flowchart of general practitioners’ participation, and dropout. **Abbreviation:** WS, workshop.
Characterization of study participants
Participants' characteristics are shown in Table 1. In most respects, they resembled the population of GPs in Hessen. The IG and CG did not differ significantly, but in the IG, there were slightly more GPs with Internet search experience and a DocCheck™ password (access to content provided by the pharmaceutical industry).

Acceptance
GPs rated the Continuing Medical Education WS as very positive (Figure 4). Over 80% of respondents rated all items as good or very good. We obtained similar ratings referring to WS organization and educational process (data not shown). At the end of the WS, most participants (78% in the IG and 69% in the CG) made a commitment to change their information management according to what they had learnt during the WS (Figure 5).

Literature search experience
In order to find answers to standardized questions, the IG used more sources per topic and spent more time searching than the CG (Table 2). Members of the IG were slightly more satisfied with the quality and amount of information as well as the time spent. Ratings for the sources’ trustworthiness and relevance were almost identical in both study arms.

Literature search sources
IG and CG preferred different sources to search literature (Table 3). The IG used more Internet sources (+22%) and targeted web sites (+37%). Overall, the WS recommendations on literature review were implemented well (67%, odds ratio 9.6, 95% confidence interval 2.6–36.4; P=0.0008).

SAC
The improvement in answers to clinical questions was the main outcome measure. In both study arms, GPs performed better with a thorough search than at baseline (Table 4). Median improvement was 11% in the CG (Wilcoxon test: P<0.0003) and 15% in the IG (Wilcoxon test: P=0.007). Although IG GPs improved slightly more than controls, the between-group comparison was statistically not significant (P=0.29).

<table>
<thead>
<tr>
<th>Group characteristics</th>
<th>External comparison: general practitioners</th>
<th>Internal comparison: complete data*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gb</td>
<td>Hc</td>
</tr>
<tr>
<td>Personal preconditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years, SD),</td>
<td>52b</td>
<td>52</td>
</tr>
<tr>
<td>Sex (female, mean %)</td>
<td>38d</td>
<td>36</td>
</tr>
<tr>
<td>Professional preconditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years in practice, (median, SD)</td>
<td>32b</td>
<td>14 (9.3)</td>
</tr>
<tr>
<td>Working in group practice (mean %)</td>
<td></td>
<td>42</td>
</tr>
<tr>
<td>Internet preconditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet access: “Yes”</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Broadband account: “Yes”</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>DocCheck™ password: “Yes”, %</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Yes: Internet search, %</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Skills* (good or very good, %)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Computer</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>English language</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Notes: *Complete data: 1: participation and evaluation of both workshops; 2: documentation and evaluation of seven literature searches; 3: answering 39 questions twice.
Gb = published statistics of the Federal Chamber of Physicians in Germany; Hc = details of the Regional Chamber of Physicians in Hessen (Landesärztekammer Hessen); IG = representative sample of the Commonwealth Fund; the DocCheck™ password allows licensee to access drug formulary and labeling information; participants have conducted Internet search for medical topics in the previous year; skills as judged by participants themselves. Scale: 1 (low) to 5 (high). *P-values for between-group comparison.

Abbreviations: IG, intervention group; CG, control group; SD, standard deviation.
Figure 4 Participants’ workshop evaluation.
Notes: (N=51, 1=very good, 6=deficient). Please note that GPs in the control group are included in this process evaluation.
Abbreviations: IG, intervention group; GPs, general practitioners; CG, control group.

Figure 5 Commitment to change at the end of the workshop (n=48).
Note: Please note that GPs in the control group are included in this process evaluation.
Abbreviation: GPs, general practitioners.

Discussion

Our study shows good WS acceptance, strong learning effects, positive search experience, and significant increase of SAC in the pre–post test comparison. Also, SAC improved more in the IG than in the CG although the difference was not statistically significant in the RCT.

The potential of online evidence in order to improve SAC skills of GPs has been demonstrated in controlled laboratory settings. Previous pre–post trials have shown an improvement from 10% to 18% in GPs’ answers to typical clinical scenarios following the use of online evidence.5,25,27 Our results correspond to these findings, with the proportion of correct answers increasing by 15% in the IG (Table 4).

The learning effect on the IG could be shown through the implementation of suggestions given in the WS (Table 3, odds ratio 9.6) of using targeted Internet sources (odds ratio 4.8),
Table 2 Literature search experience

<table>
<thead>
<tr>
<th>Group allocation</th>
<th>IG (n=20)</th>
<th>CG (n=31)</th>
<th>Difference*</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of sources used per topic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>1.9 (0.6)</td>
<td>1.4 (0.5)</td>
<td>36%</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Duration of search per topic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median min (IQR)</td>
<td>20.9 (9.6)</td>
<td>17.8 (11.5)</td>
<td>17%</td>
<td>0.34</td>
</tr>
<tr>
<td>“Yes, I am satisfied with [...]”</td>
<td>65%</td>
<td>54%</td>
<td>11%</td>
<td>0.08</td>
</tr>
<tr>
<td>“[...] quality of information”</td>
<td>80%</td>
<td>65%</td>
<td>15%</td>
<td>0.35</td>
</tr>
<tr>
<td>“[...] time spent”</td>
<td>53%</td>
<td>47%</td>
<td>6%</td>
<td>0.22</td>
</tr>
<tr>
<td>Surprised by result</td>
<td>26%</td>
<td>35%</td>
<td>9%</td>
<td>0.28</td>
</tr>
<tr>
<td><strong>Trustworthiness of sources</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (score: 1=low to 5=high) (IQR)</td>
<td>3.7 (0.6)</td>
<td>3.6 (0.7)</td>
<td>0.1</td>
<td>0.47</td>
</tr>
<tr>
<td>Results relevant for daily practice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (score: 1=low to 5=high) (IQR)</td>
<td>3.1 (0.9)</td>
<td>3.1 (0.8)</td>
<td>0.0</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Notes: *P-value for between-group comparison: Mann–Whitney U test. **Please note that physicians in the control group provided these evaluations before they were offered workshop participation. †Difference in % between IG and CG.

Abbreviations: IG, intervention group; CG, control group; IQR, interquartile range.

Table 3 Sources of information used for literature review

<table>
<thead>
<tr>
<th>Sources of information</th>
<th>IG (n=20)</th>
<th>CG (n=31)</th>
<th>Difference* OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sources per topic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Internet sources</td>
<td>80.7%</td>
<td>59.4%</td>
<td>21.3%</td>
</tr>
<tr>
<td>(n=382)</td>
<td>10.2%</td>
<td>26.2%</td>
<td>–16.0%</td>
</tr>
<tr>
<td>(of which n=119 search engines: Google, Yahoo, etc.)</td>
<td></td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td>2. Print media</td>
<td>15.7%</td>
<td>30.5%</td>
<td>–14.8%</td>
</tr>
<tr>
<td>(magazines, textbooks)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Personal interview</td>
<td>3.6%</td>
<td>10.1%</td>
<td>–6.5%</td>
</tr>
</tbody>
</table>

Internet sources

| Number of sources per topic |         |           |                |
| 1. Pharma-critical drug bulletins | 17.1% | 8.0% | 9.1% |
| (n=263, without search engines) | | | 2.9 |
| 2. Guidelines (national and international) | 38.7% | 23.0% | 15.7% |
| 3. Commercial homepages (pharmaceutical, publishing houses, clinics) | 1.8% | 31.0% | –29.2% |

Adherence to workshop recommendations

<table>
<thead>
<tr>
<th>IG (n=20)</th>
<th>CG (n=31)</th>
<th>Difference*</th>
</tr>
</thead>
<tbody>
<tr>
<td>67%</td>
<td>15%</td>
<td>52%</td>
</tr>
</tbody>
</table>

Notes: *Please note that physicians in the control group provided these evaluations before they were offered workshop participation. †Difference in % between IG and CG.

Abbreviation: OR, odds ratio.

We encouraged our participants to use targeted sources they know well despite their limitations of scope and possible bias because accessing the Internet via unspecific search engines such as Google or Yahoo is prone to error.28,29 In many areas requiring decision making under time pressure and with limited information, fast and frugal heuristics have been shown to be valid and helpful.30 Our three-step heuristic (Figure 2) not only provides rules for search and appraisal of validity but also rules for when to stop a search to save time. This seems important because lack of time is one of the main reasons why questions remain unanswered.31 Thus, our recruitment quota of 12% actually indicates that GPs have a relatively large interest in the topic of our study. The positive comments by participants are in line with this idea.

GPs in the CG had more previous methodological knowledge than we had expected. They more often obtained their information from the Internet (59%) than from printed media (31%) or through consultation of colleagues (10%). Literature has so far listed the relative importance of these sources of information in the reverse order.4,10,32 This may be an indication of a powerful CG.

Strengths and limitations of the study

The main strengths are the RCT study design and the inclusion of experienced GPs in private practice because they are often underrepresented in studies of appraisal skills training.33–35 Only 46% (Figure 3) of those approached actually terminated the study completely. Although this limits the external validity of our findings, it reflects the reality of continuing medical education where physicians are usually free to decide which program they want to take part in. Though, study
participants’ characteristics were similar to the population of GPs in Hessen and Germany.

More serious is the high dropout rate (54%) which differed between study arms (IG: 62% vs CG: 47%). As a result, the power of the study was limited and our findings therefore failed to reach statistical significance.

Apart from a lack of power due to the small sample size, a ceiling effect remains as a possible explanation for the negative result. Theoretically, even control physicians may operate at such a high level regarding their SAC that a WS is not possible for a group of GPs. We are currently working on optimizing our material for the adoption of a blended learning LP. The initiative required of each participant in a blended learning scenario could help us manage the natural process of self-selection in a constructive way.

General conclusions regarding educational projects: an RCT is an important tool to test effectiveness. Applying only a pre–post test without a CG can easily lead to an overestimation of the effectiveness of an intervention.

**Acknowledgments**

We wish to thank 137 GPs from the pilot and intervention phase who gave their time to participate in this study. We thank GPs in Bielefeld Giessen, Kassel, Marburg, and Siegen.

We thank Nicole Burchardi, PhD, Clinical Trials Unit at the University of Marburg for statistical advice. We also want to thank experts in EbM for their support in the development of the study questionnaire: Lise Bjerre, MD (Department of Family Medicine, Georg August University of Göttingen, Germany); Judith Günther, MSc (pharmfacts, Freiburg, Germany); Ralf Karger, MD, MSc (Department of Haematology, Philipps-University of Marburg, Germany); Monika Lelgemann, MD, MSc (HTA-Centre, Bremen, Germany); Dagmar Lühmann, PhD (Department of Social Medicine, University of Schleswig-Holstein, Campus Lübeck, Germany); and Antje Timmer, MD (German Cochrane Centre, University of Freiburg, Germany).

The project was funded by the Federal Ministry of Education and Research in Germany (BMBF).

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


