Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/110627

Comparing patient characteristics, type of intervention, control, and outcome (PICO) queries with unguided searching: a randomized controlled crossover trial ▣C

Arjen Hoogendam*; Pieter F. de Vries Robbé, PhD; A. John P. M. Overbeke, PhD

See end of article for authors' affiliations.

Background: Translating a question into a query using patient characteristics, type of intervention, control, and outcome (PICO) should help answer therapeutic questions in PubMed searches. The authors performed a randomized crossover trial to determine whether the PICO format was useful for quick searches of PubMed.

Methods: Twenty-two residents and specialists working at the Radboud University Nijmegen Medical Centre were trained in formulating PICO queries and then presented with a randomized set of questions derived from Cochrane reviews. They were asked to use the best query possible in a five-minute search, using standard and PICO queries. Recall and precision were calculated for both standard and PICO queries.

Results: Twenty-two physicians created 434 queries using both techniques. Average precision was 4.02% for standard queries and 3.44% for PICO queries (difference nonsignificant, $t(21)=-0.56$, $P=0.58$). Average recall was 12.27% for standard queries and 13.62% for PICO queries (difference nonsignificant, $t(21)=-0.76$, $P=0.46$).

Conclusions: PICO queries do not result in better recall or precision in time-limited searches. Standard queries containing enough detail are sufficient for quick searches.

INTRODUCTION

Quick searches for information on the Internet are becoming more and more important, but the vast amount of information available online can make it difficult to locate relevant material. Physicians want to find answers quickly and typically search less than ten minutes to find an answer to a clinical question [1–4]. Because of time constraints, physicians often prefer to review a handful of reliable sources of information rather than try to locate all the available medical evidence.

Patient characteristics, type of intervention, control, and outcome (PICO) is one of the methods that have been suggested to improve physician searches of the clinical literature. In this method, the physician is instructed to define the clinical question in terms of PICO so that the clinical question can be matched to relevant scientific literature, thereby improving retrieval. The theoretical background underlying PICO and the problems physicians face when using this method have been described by Huang [5]. One uncontrolled trial of PICO using the normal PubMed interface, a PICO interface, and a PICO interface combined with the Clinical Queries interface has been reported, but it was insufficiently powered to detect any differences between the interfaces [6]. Although the evidence behind the PICO method is still very

Highlights

- Formulation of patient characteristics, type of intervention, control, and outcome (PICO) queries does not result in a significantly higher recall or precision in PubMed when compared with standard queries.
- Queries without operators or wildcards result in similar recall and precision for PICO queries.
- Increasing knowledge of a subject area may have a stronger effect on recall than on precision.

Implications

- Teaching evidence-based quick searching techniques should focus on use of sufficient adequate terms describing the question, instead of the PICO method or the use of operators.
- Medical literature databases should contain indexing terms relevant to clinical questions to better match medical questions.
- More research is needed to explain differences in recall and precision between participants and influence of topic knowledge on recall and precision.

limited, it is recommended in evidence-based searching handbooks as a method to improve clinical queries [7, 8].
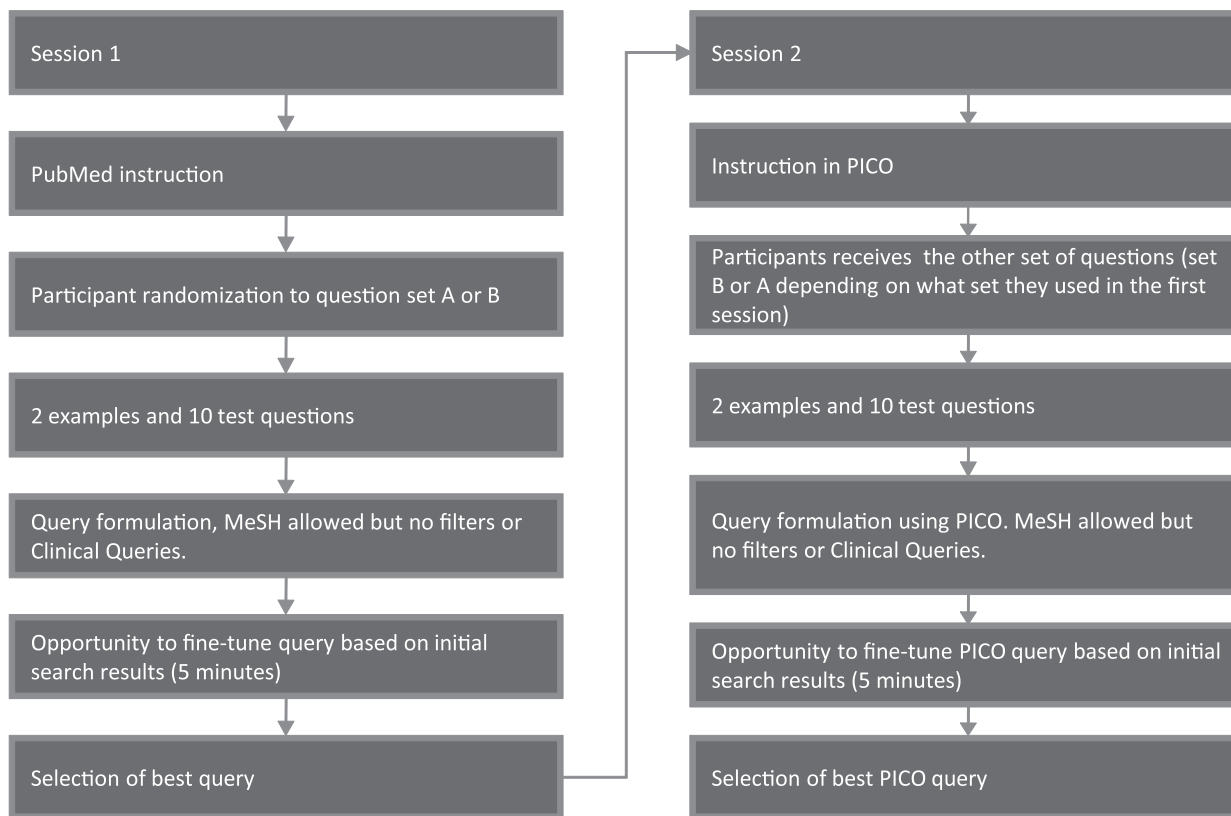
The study reported here sought to ascertain whether structuring clinical queries in the form of a PICO query, in time-restricted searches, would improve search results. A randomized crossover trial of time-restricted searches compared precision, the fraction of the retrieved documents that were relevant to the question, and recall, the fraction of the

▣C Supplemental Table 1 is available with the online version of this journal.

Figure 1
Study protocol

Session 1
PubMed instruction
Participant randomization to question set A or B
2 examples and 10 test questions
Query formulation, MeSH allowed but no filters or Clinical Queries.
Opportunity to fine-tune query based on initial search results (5 minutes)
Selection of best query

Session 2
Instruction in PICO
Participants receives the other set of questions (set B or A depending on what set they used in the first session)
2 examples and 10 test questions
Query formulation using PICO. MeSH allowed but no filters or Clinical Queries.
Opportunity to fine-tune PICO query based on initial search results (5 minutes)
Selection of best PICO query

documents that were relevant to the query that were successfully retrieved, of PICO-structured queries with unguided searches.

METHODS

The study was designed as a crossover randomized trial. Specialists from the vascular medicine staff and residents in internal medicine from the Radboud University Nijmegen Medical Centre were invited to participate in the study. All participants were familiar with searching PubMed.

Study protocol

After agreeing to the study, participants were entered in the study protocol (Figure 1) and invited to a one-hour lecture by an expert searcher explaining the basics of PubMed to ensure a basic knowledge of PubMed functionality (details, filters, history, Medical Subject Headings [MeSH], and Clinical Queries). After this explanation, they were presented with twelve therapeutic questions (two example and ten test questions in random sequence) regarding vascular medicine and asked to find a set of articles in PubMed that was both as small as possible and

contained as many useful articles as possible, judging solely from the abstract or the article title and bibliographic data. They were allowed to use MeSH but were not allowed to use Clinical Queries or other filters. After five minutes of searching, PubMed closed automatically, and the participant was asked to record, by copying and pasting, the query that delivered the most relevant articles in the smallest set of articles. Total time for the explanation and the test was a little more than two hours; time varied depending on how long it took a participant to choose the best query.

After two weeks, a second session took place. During this session, the use of PICO was explained by an expert searcher. Following instruction in the PICO method, the participants were presented with twelve different therapeutic questions (two example and ten test questions in random sequence). Below each question were four boxes representing: patient, intervention, control, and outcome. After participants filled in these boxes, the query was concatenated with the four categories surrounded by brackets. The participants were allowed to modify the PICOs if they wished, either by changing the content of a category (changing the patient or intervention by adding or removing terms) or removing a category.

Not all PICO categories had to be used; for example, if no control group could be defined, it could be left out.

Question selection and randomization

Twenty systematic reviews dealing with vascular medicine that provided references to more than five articles available in PubMed were selected from the Cochrane database. The topics of the reviews were translated to clinical questions by the authors (Table 1, online only).

This translation was then checked by a librarian to ensure that the topic of the review was reflected by the question. Four additional vascular medicine reviews that did not give more than five references to PubMed articles were used as example questions. The same number of question sets as participants was created. The order of the questions was varied in every set to ensure that all questions were evenly distributed between the first and second session and were evenly distributed across test sets. The sets were numbered and randomly assigned to a participant by a number generator.

Power

The previous study on use of PICO for clinical questions by Schardt et al. reported an average precision of 8% in unguided (non-PICO) searches [6]. The current study considered an improvement of at least 10% precision as the minimal percentage that would justify the use of the PICO method for quick searches. To detect a difference of 10% in precision with a power of 0.8 with an alpha 0.05, 177 search topics were required in each of the guided and unguided sessions. The study thus sought to obtain 20 participants who would search for answers to 10 questions in each session (200 total in each session).

Statistical analysis

The references available in PubMed for each Cochrane review were used as the gold standard containing all relevant articles to the topic. As the queries were performed on a later date than the original review, articles retrieved by the query that dated later than publication of the review were removed from the query result. The number of articles that were retrieved by the time-corrected query that were also available as references in the review was considered relevant to the question. Precision was calculated by dividing this number with the total number of articles retrieved by the time-corrected query. Recall was calculated by dividing this number with the total number of relevant articles as stated in the reference list of the review available in PubMed. SPSS version 17.0 was used to determine the standard error of the mean for recall and precision. The mean precision and recall were calculated per participant. A paired t-test was used to detect whether observed differences in recall and precision between standard and PICO sessions reached statistical difference.

RESULTS

Of 30 invited specialists and residents with interest in vascular medicine, 24 agreed to participate in the study. Eleven participants were female, 13 male; 15 were residents; 9 were specialist in internal medicine (3 fellows, 6 with a subspecialty in vascular medicine). Two physicians (1 male internist and 1 male resident) were not able to attend the second session due to causes not related to the study. Both were excluded from the analysis of results. The 22 remaining participants answered 440 questions. In 6 questions, the best query entered did not contain a search term due to errors in copying and pasting, and those searches were, therefore, excluded from analysis. The remaining 434 best standard and PICO queries were analyzed. The average precision (percentage of retrieved relevant articles of the total number of articles retrieved) was 4.02% (SE of mean=0.7%) for the standard queries and 3.44% (SE of mean=0.6%) for the PICO queries. A paired samples t-test showed that the difference in precision of −0.57% was not significant ($t(21)=-0.56$, $P=0.58$). The average recall (percentage of relevant documents that are successfully retrieved) was 12.27% (SE of mean=14.60%) for the standard queries and 13.62% (SE of mean=14.70%) for the PICO queries. A paired samples t-test showed that the difference in recall of 1.36% was not significant ($t(21)=-0.76$, $P=0.46$).

The average recall of participants showed a large variation (1.7% to 29.1%), showing that there was a large variability between participants in their ability to retrieve relevant information. Average precision was generally lower and exhibited less variability. Searches by residents had an average precision of 3.5% (SD=10.0%) and recall of 11.9% (SD=10.8%). Searches by the 3 fellows in vascular medicine had a precision of 5.4% (SD=13.8%) and recall of 13.9% (SD=21.4%), while searches by the 6 specialists in vascular medicine had an average precision of 3.3% (SD=5.8%) and recall of 15.3% (SD=23.16%). The precision and recall in relation to term count (number of terms used in query excluding operators) and the use of special operators is shown in Table 2.

Almost all queries contained more than five terms. There was no clear relation between number of search terms and recall or precision. The use of special operators also had no significant effect on recall or precision. If the PICO query retrieved too few results, participants had been advised during training to expand a search category by adding terms connected with the ''OR'' operator. Term count and the use of the ''OR'' operator was substantially higher in the PICO searches, reflecting the fact that participants used more terms to broaden the categories.

DISCUSSION

Results of this study indicate that there is no significant difference in recall or precision with PICO-guided searches in comparison with unguided searches, when performing quick (five-minute) searches for therapeutic

Table 2
Mean recall and precision of searches by term count (number of terms used in query excluding operators), use of wildcard (asterisk), and "NOT" and "OR" operators*

| | Patient characteristics, type of intervention, control, and outcome (PICO) | | | | | Standard | | | | |
| | Recall | | Precision | | Number of queries | Recall | | Precision | | Number of queries |
	Mean %	(SD)	Mean %	(SD)		Mean %	(SD)	Mean %	(SD)	
Term count										
1–4	12.50	(NA)	0.91	(NA)	1	4.17	(7.22)	0.95	(1.65)	3
5	23.13	(38.37)	2.33	(2.72)	4	14.27	(21.61)	3.34	(5.04)	23
6	17.38	(26.79)	1.85	(4.08)	11	16.43	(26.36)	4.12	(14.53)	49
7	12.12	(24.61)	6.75	(23.40)	18	8.57	(18.80)	2.62	(5.39)	33
8	8.42	(15.24)	2.05	(2.82)	18	9.68	(17.19)	3.50	(6.72)	36
9	11.69	(19.38)	2.62	(5.55)	27	7.26	(14.86)	6.60	(19.26)	27
10	6.46	(11.08)	4.05	(9.48)	30	18.31	(24.87)	2.75	(3.69)	20
10–11	16.70	(25.34)	2.38	(3.92)	38	10.73	(18.45)	7.11	(7.83)	16
12–13	18.71	(23.18)	2.96	(2.70)	20	6.48	(12.46)	1.06	(1.98)	5
14–16	15.85	(26.03)	3.98	(8.08)	27	22.60	(35.68)	5.03	(5.92)	4
17+	14.96	(18.10)	4.74	(7.76)	23	28.57	(NA)	5.00	(NA)	1
Use of wildcard (asterisk)										
None	14.01	(22.13)	3.16	(6.28)	176	12.78	(22.04)	4.02	(11.13)	192
1 or more	11.98	(19.89)	4.64	(15.97)	41	8.38	(11.80)	4.01	(6.58)	25
Use of "OR" operator										
None	11.54	(20.70)	3.36	(8.00)	56	10.72	(19.83)	4.08	(11.36)	188
1 or more	14.35	(22.05)	3.47	(9.23)	161	22.33	(26.46)	3.61	(4.45)	29
Use of "NOT" operator										
None	12.08	(21.15)	3.67	(10.23)	153	12.13	(21.15)	4.36	(11.72)	170
1 or more	17.32	(22.69)	2.90	(4.44)	64	12.78	(21.33)	2.78	(5.47)	47
Use of wildcard, "OR," or "NOT" operator										
None	9.18	(18.61)	3.18	(8.38)	42	11.51	(20.94)	4.59	(12.80)	139
1 or more	14.69	(22.29)	3.51	(9.06)	175	13.62	(21.56)	2.99	(5.05)	78

* "AND" operator excluded because PubMed treats spaces as the "AND" operator.

questions. This finding supports the one previous study on this topic, which was insufficiently powered to prove a difference. This finding may have a consequences for teaching evidence-based medicine, as the PICO method is a major component of teaching evidence-based searching [7, 8]. One reason that the PICO queries do not perform better is that article abstracts, titles, and indexing terms do not always contain the information that the PICO query is designed to retrieve. The abstract of the article, "Moderate Dietary Sodium Restriction Added to Angiotensin Converting Enzyme Inhibition Compared with Dual Blockade in Lowering Proteinuria and Blood Pressure: Randomised Controlled Trial" [9], indexed in PubMed, mentions selection of candidates as follows: fifty-two patients with nondiabetic nephropathy. In practice, the researchers selected patients with nondiabetic nephropathy and proteinuria more than one gram per day despite maximal lisinopril dosage. This selection of patients cannot be derived from the title, the abstract, or the MeSH terms added to the article. As explained by Huang, it is not always possible to translate a question into an adequate PICO, containing all four categories, especially in questions that are unlikely to have been answered with case-control studies [5]. PICO queries also have a tendency to retrieve too few articles in questions that are either related to rare diseases for are very detailed. As the PICO query looks for terms related to patient, intervention, control, *and* outcome, an article missing required information in any one of these categories will be excluded. To reduce the number of relevant articles excluded, the searcher must enter multiple synonyms and MeSH terms in each category, so that each category is comprehensive.

Traditionally in PICO, the searcher is instructed to create a query designed to locate patient information and then refine it based on the results obtained. The same procedure is then repeated for intervention, control, and finally outcome. The final combined query should then yield the optimal result. A time limit of five minutes, however, is not sufficient for such a process and perhaps explains why building a PICO query in time-limited searches did not yield better results than standard queries containing relevant terms. On the other hand, the PICO method is still useful as it emphasizes the fact that formulating an adequate question and translating the question to a query that matches the literature is crucial for finding an adequate answer. One way to improve the yield of PICO searches could be to include indexing terms related to PICOs in PubMed and assign the terms to the categories. In the case of the aforementioned article regarding dual therapy for hypertension [9], the patient characteristics might contain "proteinuria," "lisinopril," and "nondiabetic." The outcome might contain "blood pressure lowering" and "proteinuria." This might be a very effective method to make PubMed better suitable for on-the-spot searching.

As PubMed does not sort search results based on relevance, reducing the numbers of retrieved articles by increasing precision is crucial for effective searching. One simple method to increase precision is to require the presence of more terms in each article retrieved; however, this in turn will lower recall. The question is then how detailed the queries need to be to reach the optimal balance between recall and precision. Results from a previous study analyzing queries sent to PubMed reported that searches using 4 to 5 terms resulted in retrieval of 2 to 161 articles and were most frequently followed by viewing of abstracts [10]. In the current study, nearly all queries contained 4 or more terms, and the number of terms did not have any effect on recall and precision. The use of operators also did not have a significant effect on recall and precision, supporting previous results [10]. This study thus confirms that the use of 4 to 5 relevant terms without operators is likely to retrieve sufficient relevant articles to start a PubMed search, allowing the user to refine the search based on the number of articles retrieved.

Some participants reached higher recall or precision using PICO queries, and others reached higher recall or precision using unguided queries, but differences were small and did not reach significance. Some people may still be better off using PICO queries than others. It is questionable whether the small increase in recall that can be achieved, however, is worth the effort of designing a PICO query, instead of just creating a query with five relevant terms. Whether the small difference in recall and precision observed between search techniques results from the fact that some people need more time to get familiar with the PICO query technique than others or that each search technique is suitable for a certain type of person remains to be answered. The difference between residents, fellows, and specialists in vascular medicine, although insignificant, showed a trend toward higher recall. Increasing knowledge on the subject will likely result in the use of more adequate terms and therefore will yield higher recall. The trend toward higher recall for fellows and specialists in vascular medicine may be explained by this effect. Narrowing a search to increase precision, on the other hand, requires adding relevant terms as well as excluding terms with the use of the ''NOT'' operator. The latter requires the identification of common terms in nonrelevant articles that are retrieved by a search and is, therefore, independent of knowledge on the subject. This may explain why there is no trend toward better precision in fellows and specialists. Further research is needed to confirm this hypothesis.

Limitations of results

This study has several possible flaws. Failing to show a difference between PICO queries and standard queries may be related to inadequate building of queries. It may be that the time given to participants to practice the technique of translating questions to PICO queries was too short. The fact that some queries still showed erroneous use of operators despite instruction is an indication that more practice may still be needed. As operators were used in both PICO queries and standard queries, this should, however, have affected both types of queries. Another reason for the equal results between standard and PICO queries may be that the physicians in this study already used PICO data in the searches before the instruction and forced use of PICO searching. The search strings that were used by participants in PICO queries, however, contained considerably more detailed information than the unguided queries, making this assumption very unlikely.

The recall in this study may not be the actual recall, as Cochrane reviews are very strict in the selection of articles, and more articles may be suitable to answer the question. This, however, will not affect the main conclusions of this study as it is likely to have the same effect on both the original and PICO queries.

Ethical approval

No ethical approval was needed for this study.

CONCLUSION

This randomized controlled crossover trial showed that taking time to conceive PICO queries does not result in better recall or precision in searches limited to five minutes. Standard queries containing a handful relevant terms are equally efficient for quick searches on clinical questions. Some questions may be more suitable for the PICO method than others, and some physicians may perform better using the PICO method. These differences may be a focus of future research.

ACKNOWLEDGMENTS

REFERENCES

1. Coumou HCH, Meijman FJ. How do primary care physicians seek answers to clinical questions? a literature review. J Med Lib Assoc. 2006 Jan;94(1):55–60.
2. Schwartz K, Northrup J, Israel N, Crowell K, Lauder N, Neale AV. Use of on-line evidence-based resources at the point of care. Fam Med. 2003 Apr;35(4):251–6.
3. Hoogendam A, Stalenhoef AFH, de Vries Robbé PF, Overbeke AJPM. Answers to questions posed during daily patient care are more likely to be answered by UpToDate than PubMed. J Med Internet Res. 2008;10(4):e29. DOI: http://dx.doi.org/10.2196/jmir.1012.
4. Ely JW, Osheroff JA, Ebell MH, Bergus GR, Levy BT, Chambliss ML, Evans ER. Analysis of questions asked by family doctors regarding patient care. BMJ. 1999 Aug;319(7206):358–61.
5. Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a knowledge representation for clinical questions. AMIA Annu Symp Proc. 2006:359–63.

6. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak. 2007;7:16. DOI: http://dx.doi.org/10.1186/1472-6947-7-16.

7. Straus S. Evidence-based medicine: how to practice and teach EBM. 3rd ed. Edinburgh, UK; and New York, NY: Elsevier/Churchill Livingstone; 2005.

8. Guyatt G, Evidence-Based Medicine Working Group; American Medical Association. Users' guides to the medical literature: a manual for evidence-based clinical practice. Chicago, IL: AMA Press; 2002.

9. Slagman MC, Waanders F, Hemmelder MH, Woittiez AJ, Janssen WM, Lambers Heerspink HJ, Navis G, Laverman GD, HOlland NEphrology STudy Group. Moderate dietary sodium restriction added to angiotensin converting enzyme inhibition compared with dual blockade in lowering proteinuria and blood pressure: randomised controlled trial. BMJ. 2011 2011 Jul 26;343:d4366.

10. Hoogendam A, Stalenhoef AFH, de Vries Robbé PF, Overbeke AJPM. Analysis of queries sent to PubMed at the point of care: observation of search behaviour in a medical teaching hospital. BMC Med Inform Decis Mak. 2008;8:42. DOI: http://dx.doi.org/10.1186/1472-6947-8-42.

AUTHORS' AFFILIATIONS

Arjen Hoogendam†, Arjen.Hoogendam@gmail.com; **Pieter F. de Vries Robbé, PhD,** P.deVriesRobbe@elg.umcn.nl; **A. John P. M. Overbeke, PhD,** J.Overbeke@elg.umcn.nl; Department of Medical Informatics, Radboud University Nijmegen Medical Centre, P.O. Box 9101, 6500 HB, Geert Grooteplein 21, Nijmegen, The Netherlands

† Present address of Arjen Hoogendam: Arjen.Hoogendam@gmail.com, Department of Internal Medicine, Dr. Horacio E. Oduber Hospitaal, Dr. Horacio E. Oduber Hospitaal Boulevard #1, Oranjestad, Aruba.