PICO as a Knowledge Representation for Clinical Questions

Xiaoli Huang, Jimmy Lin, Ph.D. and Dina Demner-Fushman, M.D., Ph.D.
University of Maryland, College Park

Abstract

The paradigm of evidence-based medicine (EBM) recommends that physicians formulate clinical questions in terms of the problem/population, intervention, comparison, and outcome. Together, these elements comprise a PICO frame. Although the framework was developed to facilitate formulation of clinical questions, the ability of PICO structures to represent physicians' information needs has not been empirically investigated. This paper evaluates the adequacy and suitability of PICO frames as a knowledge representation by analyzing 59 real-life questions in primary care. We discovered that only two questions in our corpus contained all four PICO elements, and that 37% of questions contained only intervention and outcome. Results revealed structural frame patterns that cluster according to the type of clinical question, i.e., therapy, diagnosis, prognosis, and etiology. We found that the PICO framework is primarily centered on therapy questions, and is less suited to representing other types of clinical information needs. Challenges in mapping natural language questions into PICO structures are also discussed. Although we point out limitations of the PICO framework, our study as a whole reaffirms its value as a tool to assist physicians practicing EBM.

Introduction

Clinicians have from 0.7 to 18.5 questions for every 10 patients cared for[1,2]. However, answers to two-thirds of the questions are either not pursued or pursued but not found[3,4]. A subsequent analysis shows that almost all unanswered questions could be answered after improved query formulation and search[2]. Therefore, helping doctors to articulate their clinical information needs through well-built, focused questions has become one of the focal points in evidence-based medicine (EBM)[5]. EBM provides an explicit framework and guidance for formulating a patient-specific clinical question[6]. According to EBM, articulating a clinical question in terms of its four anatomic parts—Problem/Population, Intervention, Comparison, and Outcome (PICO)—facilitates searching for a precise answer.

This study investigates the suitability of the PICO frame as a knowledge representation for clinical questions posed in natural language by practicing physicians. To our knowledge, no research has studied the adequacy and flexibility of the PICO representation and whether it is complete in terms of being able to capture salient characteristics of clinical questions. The present study attempts to address these issues by manually mapping real clinicians' questions into PICO frames and examining the results.

Background

A previously-explored approach to understanding the nature of clinical information needs is to collect and to classify real clinical questions from physicians. Through such analyses, studies have introduced question taxonomies in varying levels of details[7,8]. Taxonomies capable of "covering" a large fraction of clinical questions with a smaller set of "question templates" facilitate access to relevant evidence in the medical literature. Nevertheless, previous studies focus on the surface form of the questions and do not take into account EBM principles for formulating "answerable" clinical questions.

The well-built clinical question, focused and well-articulated in all four components of its anatomy, is widely believed to be the key to efficiently finding the best evidence and also the key to evidence-based decisions[6,9]. Empirical studies have shown that the use of PICO frames improves the specificity and conceptual breakdown of clinical problems[10], elicits more information during pre-search reference interview, and leads to more complex search strategies and more precise search results[11].

There are few studies that describe the usability and acceptability of PICO in general, and even less prior work on PICO application in computerized information retrieval systems. A small questionnaire-based study reported a PICO interface for handhelds considered as easy to use and useful in searching MEDLINE[12]. However, the use of PICO-structured frames does not always translate into higher satisfaction[10,11].

To better understand the adequacy and flexibility of the PICO framework as a knowledge representation, we coded a set of real-world questions asked by physicians into PICO frames. Through the mapping and subsequent analysis, we addressed the following research questions:

1. How well are real-life clinical questions structured according to PICO standards?
2. How well-suited is the PICO frame as a knowledge representation for clinical questions?

3. What concepts and relationships are not adequately captured by PICO representations?
4. Is PICO frame equally suitable for representing different types of clinical questions?

Methodology

Data Collection

We gathered 59 real-world clinical questions from two on-line sources: Family Practice Inquiries Network (FPIN)[*] and Parkhurst Exchange[†]. The question collection process was guided by typical instance sampling[13] rather than random sampling, because the goal is not to obtain a fully representative, but a typical sample of real-life clinical questions. According to the literature, approximately 33% of questions asked by clinicians are about treatment, 25% about diagnosis, and 15% about pharmacotheapeautics. Together, they account for over 70% of clinicians' questions[7,10]. Guided by this distribution, four types of clinical questions were gathered: therapy (25), diagnosis (15), prognosis (7), and etiology (12).

Coding Clinical Questions with PICO

The questions were coded into PICO frames independently by the first and the third author (with backgrounds in library science and medicine, respectively). The comparison and reconciliation of the resulting PICO representations was guided by the second author. This being an exploratory study and the first of its type that we are aware of, the primary purpose of independent coding was to preserve multiple perspectives, rather than to enforce uniformity for the sake of measuring inter-coder agreement. Therefore, no formal instructions or protocol beyond standard EBM guidelines was given to the coders.

Analysis of the Results

Our corpus of 59 questions was first evaluated for structural completeness. Based on the finding that clinical questions were less likely to go unanswered when the question identified the proposed intervention and desired outcome[14], we used the co-occurrence of intervention and outcome as an indication of the structural completeness of a question.

We then analyzed the prevalence of each PICO element. This analysis gave rise to structural frame patterns that represented prototypical therapy, diagnosis, etiology, and prognosis questions. In addition, semantic classes of concepts present in the 59 clinical questions were identified. This allowed us to construct the typical mapping relationships between semantic entities and PICO elements.

Finally, challenges encountered during the process of coding these clinical questions were gathered and categorized into emergent themes. This yields a qualitative evaluation of the adequacy of PICO as a knowledge representation for clinical questions.

Results

Structural Completeness of Clinical Questions

In our collected corpus, only two out of 59 questions specified all four PICO elements and 37.3% of questions contain only intervention and outcome. Table 1 provides an overview of how often different PICO elements are presented in each question type.

Independent of question type, problem/population and intervention are the most frequently addressed PICO elements (50 and 49 out of 59 respectively), followed by population (29 out of 59), then by outcome (27 out of 59). In contrast, comparison is rarely mentioned (only 3 out of 59).

Table 1. Structural Completeness for Four Types of Clinical Questions. 1–number of questions; 2–all elements present; 3–intervention and outcome present

	Therapy	Diagnosis	Etiology	Prognosis	Total
1	25	15	12	7	59
2	1 (4.0%)	1 (6.7%)	0 (0%)	0 (0%)	2 (3.4%)
3	16 (64.0%)	5 (33.3%)	0 (0%)	1 (14.3%)	22 (37.3%)

Prototypical PICO Representations

Manual mapping of clinical questions into PICO representations allows us to derive prototypical query patterns that capture prevalent structural themes.

Table 2. Structural patterns and examples for 25 therapy questions. P–Problem/Population; I–Intervention; C–Comparison; O–Outcome

[P][I][O**?**] (10)	Could stimulants be useful for chronic fatigue syndrome?
[P][I**?**] (8)	What is the best treatment for analgesic rebound headaches?
[I][O**?**] (2)	What protective effects do vitamins E, C, and beta carotene have on the cardiovascular system?
[P][I**?**][O] (2)	What regimens eradicate Helicobacter pylori?
[P][I][C][O**?**] (2)	Do acetaminophen and an NSAID combined relieve osteoarthritis pain better than either alone?
[I**?**] (1)	What is the most effective nicotine replacement therapy?

[*] http://www.fpin.org/

[†] http://www.parkhurstexchange.com/

All of the 25 therapy questions contained an identifiable intervention. All but two therapy questions describe the problem, the population, or both. Overall, 64% of the questions provide explicit statements of desired outcomes. Structural patterns observed in therapy questions and their frequency are shown in Table 2. A question mark notes the element questioned about, e.g., [O**?**] indicates that an outcome serves as the answer to the question.

For diagnosis questions, emphasis is placed on symptoms, which belongs in the population slot (11 of 15), hypothesized disease, which belongs in the problem slot (12 of 15), and diagnostic approach, which belongs in the intervention slot (10 of 15). Five of 15 diagnosis questions asked for actual diagnoses, another five asked about the efficacy of specific diagnostic approaches or tests, and the last five asked about diagnostic approaches or tests that can be applied to specific patient situations. Although the PICO framework collapses the two "P's" (population and problem), we discovered a need to explicitly distinguish between the two in diagnosis questions: "population" is used to describe patients' symptoms, while "problem" is used to describe the hypothesized disease. Five of 15 diagnosis questions consisted of explicit population and problem elements only, e.g., "What is the differential diagnosis of chronic diarrhea in immunocompetent patients?"

The structure of etiology questions examined in this study is homogenous. All 12 questions describe the problem and inquire about its etiology, following the pattern [P][I?], e.g., "What are the causes of hypomagnesemia?" Although counter-intuitive, causes are best captured in the intervention slot (see discussion section for more detail).

Table 3. Structural patterns and examples of the 7 prognosis questions. P1–Problem; P2–Population; I–Intervention; O–Outcome

[P1][O**?**] (5) What is the prognosis for acute low back pain?
[P1][P2][I][O**?**] (1) A patient with stable creatinine and IgA Urology after a renal biopsy. His blood pressure and proteinuria are normal while he takes his enalapril. What is his prognosis?
[P2][O**?**] (1) What is the prognosis for chronic active hepatitis, cirrhosis, and hepatoadenocarcinoma in an active asymptomatic 45-year-old man with no history of illness, strongly positive result for HBsAg and practically none for HBsAb ?

Prognosis questions focus on patient outcomes, given a diagnosed problem or a patient profile (population). Various structural patterns are shown in Table 3.

Mapping of Semantic Classes

To investigate how specific semantic classes relate to PICO elements, we manually grouped concepts into semantically-related categories, which loosely correspond to UMLS[15] semantic types:

Problem
- [DISEASE], e.g., "panic disorder"
- [BEHAVIOR], e.g., "oppositional behaviors"
- [SYMPTOM], e.g., "leg cramps"

Population
- [AGE], e.g., "40-year- old"
- [GENDER], e.g., "male"
- [TREATMENT STATUS], e.g., "delayed treatment"
- [PHYSICAL CONDITION], e.g., "healthy"
- [MEDICAL HISTORY], e.g., "with prior attacks"
- [TREATMENT & DRUG], e.g., "taking hormone replacement therapy"
- [DISEASE], e.g., "nonvalvular atrial fibrillation"
- [SYMPTOM], e.g., "chronic cough"

Intervention & Comparison
- [TREATMENT & DRUG], e.g., "warfarin"
- [PROCEDURE], e.g., "transvaginal ultrasound"
- [DIAGNOSTIC TEST], e.g., "Pap smear"
- [EXPOSURE], e.g., "maternal smoking"
- [DISEASE], e.g., "a flare-up of the Crohn's"
- [SYMPTOM], e.g., "a very low serum iron"

Outcome
- [TREATMENT OUTCOME], e.g., "fibroid volume reduction"
- [PATIENT OUTCOME], e.g., "decreased mortality"

We noted that some semantic classes show strong, predictable associations to specific PICO elements. For example, [AGE], [GENDER], [TREATMENT STATUS], [PHYSICAL CONDITION], and [MEDICAL HISTORY] are always mapped to the population slot. On the other hand, there are also semantic classes that can be mapped to more than one PICO element. Semantic classes such as [TREATMENT & DRUG], [DISEASE], and [SYMPTOM] are likely to take different roles for different question types. For example, [TREATMENT & DRUG] is considered an intervention in the context of a therapy question, but may be part of the population in a prognosis question, i.e., a woman on hormone replacement therapy. We believe that potential confusion in the associations between semantic classes and PICO elements presents a potential barrier to the formulation of clear clinical questions.

Discussion

As shown in Table 1, only 22 of 59 questions in the study contain both the intervention and outcome elements. This confirms the findings of Bergus et al.[14] who discovered that few real-life clinical questions meet the minimum structural requirements for facilitating precise searches. Therapy questions (64%) are most likely to be structured with both intervention and outcome, followed by diagnosis questions (33.3%). Prognosis questions (14.3%) and etiology questions were least structured (0%).

Challenges in Structuring Clinical Questions

Our study revealed a number of challenges in applying the PICO framework to analyzing clinical questions. We describe our observations below:

Inability to reconstruct the original question. Given a PICO frame, can we recover the original clinical question? For example, does the frame [*Problem:* hypomagnesemia, *Intervention:* **?**] correspond to "What is the most effective treatment for hypomagnesemia?" or "What are causes of hypomagnesemia?" This ambiguity, however, is easily resolved if the clinical task, e.g., therapy or etiology, is known. However, this suggests that the clinical task is an essential component of PICO representations, which would require a minor modification of the existing framework.

Inability to encode fine-grained relationship between frame elements. Consider the following question:

> Is there any evidence to show that selective serotonin reuptake inhibitor (SSRI) use carries a risk of impulsive suicidal or homicidal behaviour, or is it just a case of association, in that those most likely to perform such acts are also most likely on mood-stabilizing medications for their underlying psychopathology?

It is difficult to represent the above question in a PICO frame without losing the fine-grained semantic relationships between concepts. The PICO framework mainly relies on inherent semantic relationships between concepts to connect different elements. For example, with etiology questions, the connection between interventions and problems is assumed to be causal. As a result, the PICO frame is ill-suited to questions that challenge these implicit relations.

No explicit temporal/state model. The PICO frame describes the state of affairs at a frozen point in time. However, temporal progress is a salient element of many clinical questions[16], and temporal concepts are often critical to retrieving precise results.

For "medication states", we can work around this problem by interpreting it as a part of patient profile, i.e., population, as in the following question:

> What is the interval for monitoring warfarin therapy once therapeutic levels are achieved?
> *Population:* therapeutic levels are achieved

Consider another common use of temporal concepts, as illustrated with the question: "Are there any advances in the treatment of motion sickness since 90s?" The PICO framework contains no provisions for capturing such temporal modifications. At present, physicians must consider metadata requirements beyond the PICO frame in formulating their searches, e.g., restricting searches to specific publication dates.

Overloaded slots. We have observed certain types of clinical questions that cannot be intuitively captured by the existing PICO framework. For example, the standard PICO frame combines problem and population into a single "P" element. However, for diagnosis questions, as mentioned earlier, the most common structural pattern consists of a population and a hypothesized disease. To capture such questions, the "P" slot needs to be more finely articulated, explicitly separating problem from population. Otherwise, it would be problematic for questions like:

> How would you manage a woman with brownish discharge from one of her breasts? She is premenopausal (less than 50 years old)

Another limitation of the PICO framework occurs with etiology questions, which, in our sample, all inquire about causes of diseases. Naturally, the disease fills the problem slot. But in what slot does the cause belong? Intervention is the closest match, but this placement is highly counter-intuitive. The intervention is generally thought of as "something done" to affect the problem, as in treating a disease with a drug. The encoding of etiology questions reverse the direction of causality normally associated with other question types. The result is potential confusion in the formulation of well-specified clinical question.

Inability to Capture Anatomical Relations. The PICO frame is unable to capture anatomical relations that may be relevant in a clinical question. Nevertheless, questions focusing on particular portions of the human anatomy are quite common, for example:

> What protective effects do vitamins E, C, and beta carotene have on the cardiovascular system?

Quite simply, there isn't a slot in the PICO framework capable of capturing "body parts". Given the small size of our sampled questions, it is difficult to determine whether there are more concepts in real-life clinical questions that are not covered by the PICO frame.

Summary

Our study shows that the PICO framework is best suited for capturing therapy questions, and considerably less-suited for diagnosis, etiology, and prognosis questions. In some cases, it is difficult to encode certain question classes without modifying the existing PICO structure or introducing counter-intuitive notions. Given that the PICO framework is a well-established tool for formulating clinical queries, any limitations of the framework itself could potentially impact the quality of clinical evidence retrieved under its guidance. This study has revealed a number of challenges associated with PICO analysis, which will serve as a basis for refining the principles of clinical query formulation.

Conclusions

This study aimed at investigating the adequacy and suitability of the PICO frame as a knowledge representation for clinical questions. Our exploration focused on a manual analysis of 59 real-world clinical questions drawn from online sources. Overall, results reaffirm the PICO framework as a method for structuring clinical questions, since natural language questions were found to lack the elements that comprise a well-formed query in most cases. Nevertheless, we encountered many challenges in employing PICO frames as a representation for clinical information needs. However, a better understanding of the advantages and limitations of this framework will translate into more effective strategies for retrieving relevant clinical evidence. We hope that these insights will ultimately translate into next-generation retrieval systems that leverage computational models of evidence-based medicine to automatically answer clinical questions[17].

References

1. Ely JW, Osheroff JA, Chambliss ML, Ebell MH, Rosenbaum ME. Answering physicians' clinical questions: Obstacles and potential solutions. J Am Med Inform Assoc. 2005 Mar-Apr;12(2):217-24.

2. Gorman PN, Helfand M. Information seeking in primary care: How physicians choose which clinical questions to pursue and which to leave unanswered. Med Decis Making. 1995 Apr-Jun;15(2):113-9.

3. Chambliss ML, Conley J**.** Answering clinical questions. J Fam Pract. 1996 Aug;43 (2): 140-144.

4. Currie LM, Graham M, Allen M, Bakken S, Patel V, Cimino JJ. Clinical information needs in context: an observational study of clinicians while using a clinical information system. AMIA Annu Symp Proc. 2003;:190-4.

5. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996 Jan 13;312(7023):71-2.

6. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club. 1995 Nov-Dec;123(3):A12-3.

7. Ely JW, Osheroff JA, Gorman PN, Ebell MH, Chambliss ML, Pifer EA, Stavri PZ. A taxonomy of generic clinical questions: classification study. BMJ. 2000 Aug 12;321(7258):429-32.

8. Jerome RN, Giuse NB, Gish KW, Sathe NA, Dietrich MS. Information needs of clinical teams: analysis of questions received by the Clinical Informatics Consult Service. Bull Med Libr Assoc. 2001 Apr;89(2):177-84.

9. Ebell MH. Information at the point of care: answering clinical questions. J Am Board Fam Pract. 1999 May-Jun;12(3):225-35

10. Cheng GY. A study of clinical questions posed by hospital clinicians. J Med Libr Assoc. 2004 Oct;92(4):445-58.

11. Booth A, O'Rourke AJ, Ford NJ. Structuring the pre-search reference interview: a useful technique for handling clinical questions. Bull Med Libr Assoc. 2000 Jul;88(3):239-46.

12. Fontelo P, Nahin A, Liu F, Kim G, Ackerman MJ. Accessing MEDLINE/PubMed with Handheld Devices: Developments and New Search Portals Proc. of the 38th Hawaii International Conference on System Sciences. 2005;:1-5.

13. Lindlof TR, Taylor BC. Qualitative Communication Research Methods, 2nd Ed. Thousand Oaks, CA: Sage Publications; 2002.

14. Bergus GR, Randall CS, Sinift SD, Rosenthal DM. Does the structure of clinical questions affect the outcome of curbside consultations with specialty colleagues? Arch Fam Med. 2000 Jun;9(6):541-7.

15. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Meth. Inform. Med. 1993; 32: 281-291.

16. Green ML, Ruff TR. Why do residents fail to answer their clinical questions? A qualitative study of barriers to practicing evidence-based medicine. Acad Med. 2005 Feb;80(2):176-82.

17. Demner-Fushman D, Lin J. Knowledge extraction for clinical question answering: preliminary results. In: Proceedings of the AAAI-05 Workshop on Question Answering in Restricted Domains; 2005;Jul 9-13; Pittsburgh. p. 1-10