

# What's in an Abstract? PICOS Reporting Frequencies to Inform AI-Assisted Screening

Allie Cichewicz<sup>1</sup>, Marius Sauca<sup>1</sup>, Kevin Kallmes<sup>1</sup>  
<sup>1</sup>Nested Knowledge, St. Paul, MN

MSR127



{NESTED}  
KNOWLEDGE

## Introduction

- Large language models (LLMs) are increasingly being utilized to assist researchers with abstract screening in systematic literature review workflows.
- Since LLMs do not require task-specific training and depend on explicit inclusion/exclusion criteria provided by researchers, incomplete reporting in abstracts due to word limit requirements or ambiguous language often require inferential judgments traditionally made by experienced reviewers.
- Optimizing prompts for LLM-assisted screening requires understanding of which concepts are reliably reported in abstracts and where gaps necessitate flexible criteria or human oversight.

## Objective

To characterize the reporting frequency of key **Population, Intervention/Comparator, Outcome, and Study Design (PICOS)** elements used in abstract screening across randomized controlled trial and observational study abstracts to inform prompt design for LLM-assisted literature screening.

## Methods

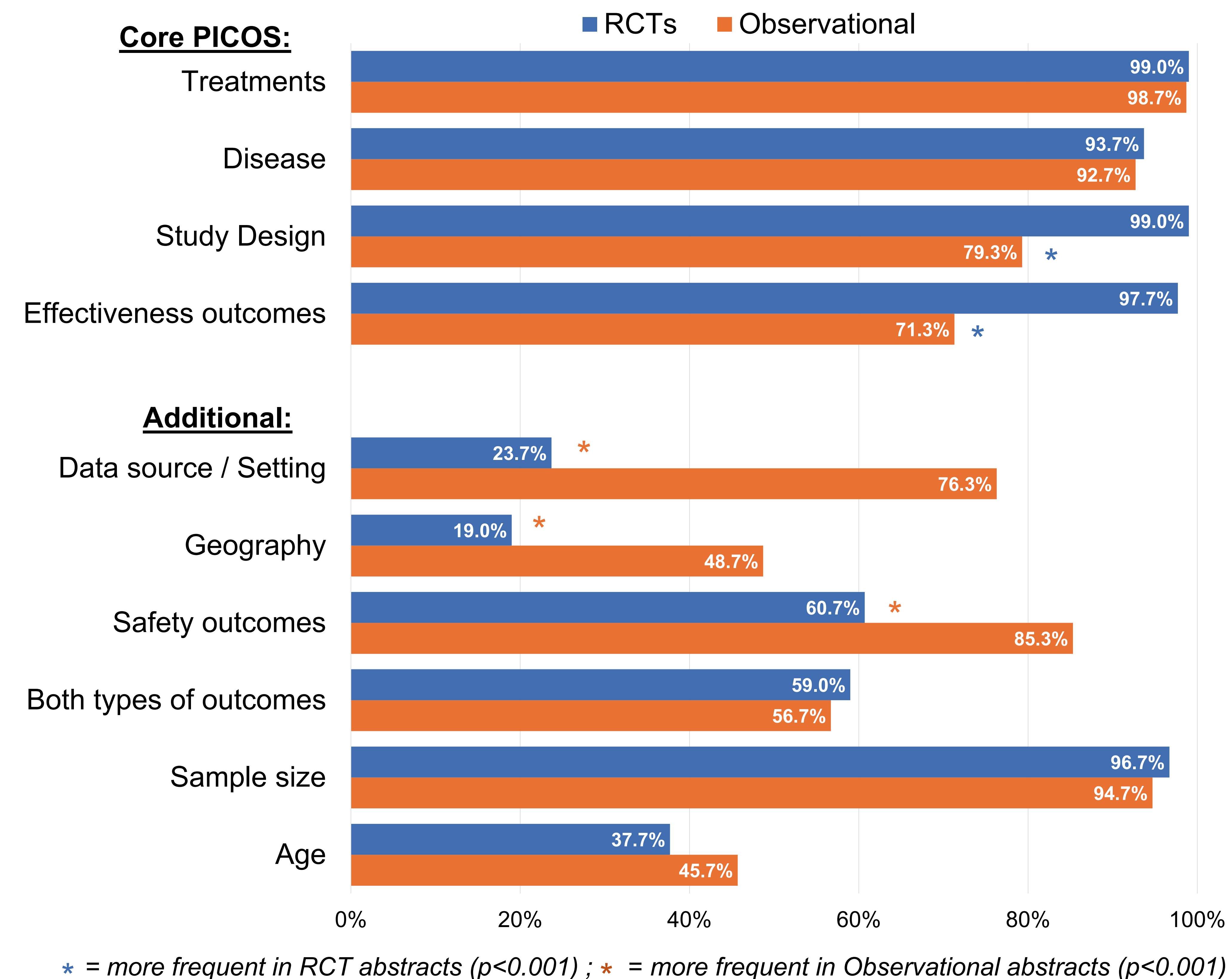
- PubMed was searched via the Nested Knowledge platform to identify a random sample of 600 abstracts (300 RCTs, 300 observational studies) evaluating treatment efficacy, effectiveness, and/or safety.
- Each abstract was reviewed for the presence of core PICOS elements (e.g., population/disease, treatments, outcomes) and additional elements that may be defined by a review protocol: age, geography, sample size, etc.
- Reporting frequencies for each PICOS element were compared between RCT abstracts and observational study abstracts using Fisher's exact test.

### Example Observational and RCT Abstract Screening Criteria:

Abstract Criteria		Abstract Criteria	
Name	Question	Name	Question
:: Study type	Does the abstract report that it is an observational stu...	:: Study Type	Does the study report that it is a randomized contro...
:: Data source	Does the study report where data or patients are bein...	:: Phase	Does the study report the trial phase?
:: Geography	Does the abstract report which country or region the st...	:: Setting	Does the study report what setting it was carried out in?
:: Disease	Does the abstract report what disease state is being a...	:: Geography	Does the study report what geography it was conducte...
:: Age	Is the age of participants reported?	:: Disease	Does the study report what disease area is being treat...
:: Treatments	Does the abstract report which interventions or treatm...	:: Sample Size	Does the study report number of participants?
:: Effectiveness/Outcomes	Does the abstract report which effectiveness outcome...	:: Age	Does the study report age of the participants?
:: Safety Outcomes	Does the abstract report which safety outcomes are b...	:: Treatments	Does the study define what the interventions and com...
:: Sample size	Is the number of participants reported?	:: Efficacy Outcomes	Does the study report efficacy outcomes?
		:: Safety Outcomes	Does the study report safety outcomes?
		:: Registration Information	Does the study report the trial registration number?

## Results

Figure 1. PICOS Elements Reporting Rates for RCT and Observational Study Abstracts



- The proportions of abstracts that included each core PICOS criteria and other screening criteria are reported in **Figure 1**.

### Core PICOS Criteria:

- Among the abstracts analyzed, treatment/intervention (RCTs: 99.0% vs. Obs: 98.7%, p=1.00), and disease/condition (93.7% vs. 92.7%, p=0.75) demonstrated consistently high reporting frequencies with no significant differences between RCTs and observational study abstracts.
- RCT abstracts more frequently reported study design (99.0% vs. 79.3%, p<0.001) and efficacy/effectiveness outcomes (97.7% vs. 71.3%, p<0.001).

### Additional Criteria:

- Observational studies more frequently reported data source/setting (Obs: 76.3% vs. RCT: 23.7%, p<0.001), geography (48.7% vs. 19.0%, p<0.001), and safety outcomes (85.3% vs. 60.7%, p<0.001).
- Age was moderately reported in both RCTs and observational studies.
- Among RCTs, registration (19.0%) and trial phase (15.3%) were infrequently reported (not applicable to measure in observational studies).

## Conclusions

- LLM-assisted screening reliably identifies core PICOS elements (treatment, disease) due to high reporting consistency, supporting binary yes/no classification when information is explicitly stated.
- However, study design-specific gaps, including underreported effectiveness outcomes in observational studies and underreported safety, setting, and geography in RCTs, mean that absent information cannot be treated as implicit "no" responses, which in turn risk inappropriate abstract exclusions. A forced binary framework therefore compounds this risk for underreported criteria.
- Incorporating an "unclear" or "not reported" category for low-frequency PICOS elements better mirrors human reviewer reasoning, preserves borderline studies, and enables transparent auditing and iterative prompt refinement.