Reporting of Study Details in Abstracts: Informative for Artificial Intelligence (AI) or Underwhelming?

Allie Cichewicz[1], Marius Sauca[1], Kevin Kallmes[1]
[1]Nested Knowledge, St. Paul, MN

MSR177

{NESTED} KNOWLEDGE

Introduction

- Advancements in AI, particularly large language models (LLMs) and retrieval systems (e.g., RAG), enable automated searching and streamline the initial review phase to help researchers quickly identify relevant studies.
- However, accuracy is limited by study information described in abstracts.
- Checklists like CONSORT-A, PRISMA-A, and STARD have helped standardize reporting, but differences still exist between authors, journals, and study types.

Objective

The aim of this analysis is to synthesize evidence on abstract reporting frequencies to identify trends and gaps to inform criteria-based screening methods leveraging AI.

Methods

- A comprehensive, living review was undertaken in Nested Knowledge to identify studies that evaluate abstract reporting and the prevalence of key methodological concepts commonly used to determine study eligibility during the title/abstract screening step of a literature review.
- PubMed was searched (n=579 records; as of December 2025), supplemented by expert recommendations (n=18).
- Records were manually screened using predefined criteria for studies evaluating abstract reporting as a primary aim that audited a sample of published studies and used a recognized checklist (e.g., CONSORT, PRISMA) as the evaluation basis
- For included studies, 11 reporting concepts were extracted. Frequency of reporting was averaged across all studies, and by study type.

Results

- 47 studies were included covering 37,177 abstracts, predominantly from randomized controlled trials (RCTs) (10,132 [27.3%]; n=33 studies), systematic reviews (742 [2.0%]; n=6), observational studies (650 [1.8%]; n=2), diagnostic accuracy (616 [1.7%]; n=4), RCT + observational (130 [0.4%]; n=1), and all study types (24,907 [67.0%]; n=1).
- Across all study types (**Figure 1**), intervention/treatment (88%) and disease/condition (86%) were consistently well-reported within abstracts, while participant eligibility (60%), efficacy/effectiveness outcomes (62%), and sample size (58%) showed moderate reporting. Safety outcomes (38%), data source/setting (38%), and registration details (27%) were poorly reported.

Results (cont'd)

Figure 1. Frequency of Reporting of Key Concepts

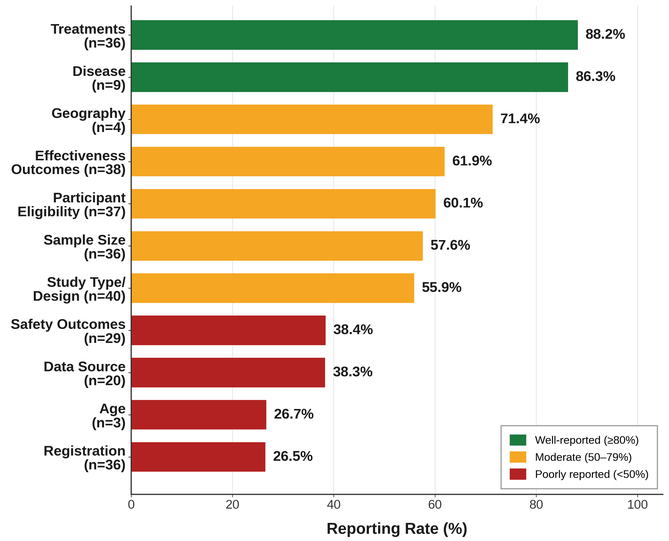
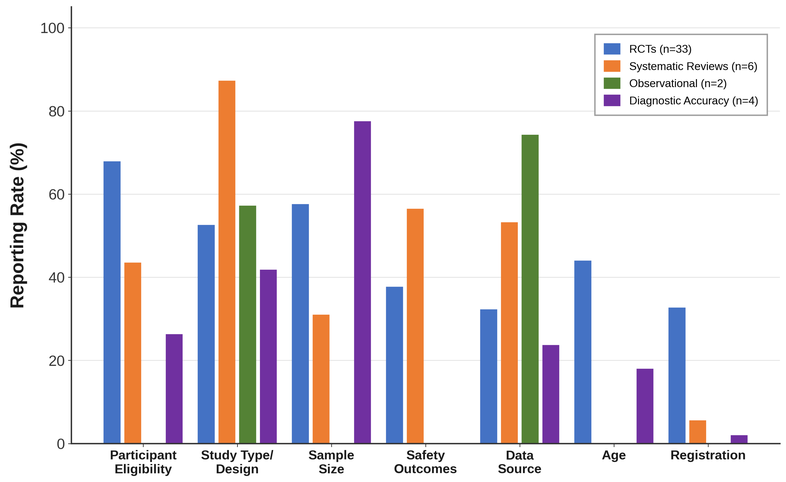


Figure 2. Frequency of Reporting by Study Type



Omitted categories ranged: treatments (79.7–99.5%), disease (84.0–100.0%), geography (58.5–100.0%), effectiveness outcomes (59.2–80.2%).
Missing bars indicate insufficient data for that study type. RCT registration information ranged from 1–99% across studies.

- Reporting patterns notably by study type (**Figure 2**):
  - **Diagnostic accuracy** studies had the strongest sample size reporting (78%) but the poorest participant eligibility (26%) and registration (2%)
  - **Systematic reviews** excelled at study type identification (87%) but had weak registration (6%) and sample size reporting (31%)
  - **RCTs** showed particularly poor data source/setting reporting (32%) with highly variable trial registration (1–99% across studies).
  - **Observational studies**, though limited in number (n=2), had the highest data source reporting (74%) but lacked sufficient data for most other elements.

Limitations

- The existing evidence base is heavily skewed toward RCTs (70%).
- Most assessments used reporting guidelines (e.g., CONSORT) requiring rigorous methodological details rather than basic concept presence, potentially overestimating gaps relevant to AI-assisted screening where partial reporting often suffices.

Conclusions

- Core clinical content (treatments, disease) is reliably reported in abstracts (>85%), but methodological details like registration, data source, and safety outcomes remain inconsistently reported across all study types.
- Future assessments focused on PICO-based concept presence rather than checklist compliance would provide more actionable insights for prompt engineering and AI-assisted screening.
- AI screening tools should account for systematic reporting gaps, particularly in registration, safety outcomes, and data source, when designing extraction criteria from abstracts.

Disclosures: Authors are contractors, employees, directors, and/or equity holders of Nested Knowledge and this study was conducted using the Nested Knowledge platform.

Presented at ISPOR 2026 - Philadelphia