

Online Supplementary Material

Evaluating AI Performance in Systematic Literature Reviews for HEOR: Methods and Recommendations. *JHEOR*. 2026;13(2):46-54. [doi:10.36469/jheor.2026.165173](https://doi.org/10.36469/jheor.2026.165173)

Figure S1: Methods Used in the Traditional SLR

Table S1: Data Extraction Process Using Smart Tags

Table S2: Data Sources

Table S3: PICOS Eligibility Criteria

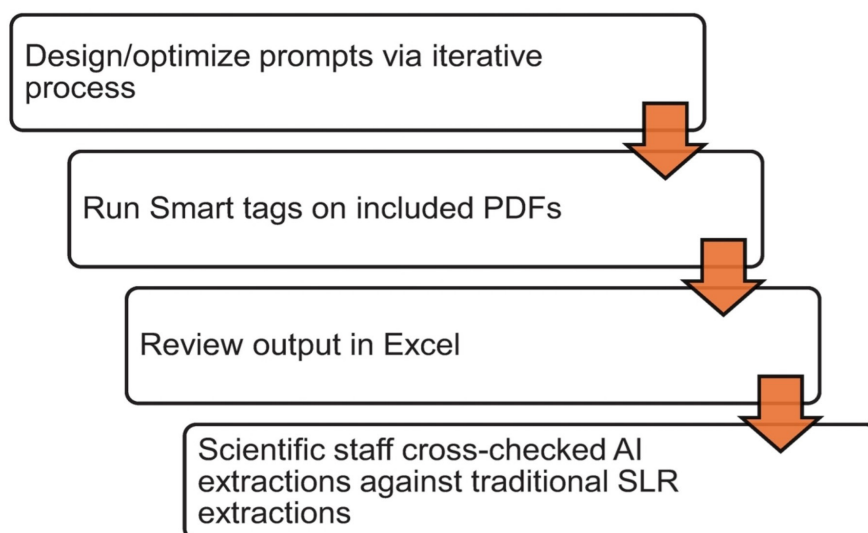
Table S4: AI Configuration

Table S5: Prompts Used for Full-Text and Data Extraction

Figure S2: Hierarchy of Prompts Designed for AI Data Extraction

This supplementary material has been provided by the authors to give readers additional information about their work.



Figure S1. Data Extraction Process Using Smart Tags

Abbreviations: AI, artificial intelligence; SLR, systematic literature review.

Table S1. Methods Used in the Traditional SLR

SLR Task	Methods
Title/abstract screening	Double screening: each record was screened against the prespecified PICOS eligibility criteria by 2 independent reviewers. Any disagreements were resolved in consultation with a senior systematic reviewer.
Full-text screening	Records that passed the title/abstract screening were obtained as full-text, including all available supplementary materials. Double screening: each full-text and supplementary materials were screened against the pre-specified PICOS eligibility criteria by 2 independent reviewers. Any disagreements were resolved in consultation with a senior systematic reviewer.
Data extraction and validation	Data were extracted by 1 reviewer and validated in full by an independent reviewer. Any disagreements were resolved in consultation with a senior systematic reviewer.

Abbreviation: PICOS, population, interventions, comparisons, outcomes, and study design.

Table S2. Data Sources

Types of Sources	Data Source via OvidSP
Literature databases	Embase MEDLINE and MEDLINE In-Process EconLit
Conference proceedings	Conference abstracts were searched via the searches in Embase. The in-house SLR conducted additional hand searches on the websites of the conferences listed below. The AI SLR included only conferences indexed in Embase (years indexed in Embase are shown) AACR: 2020, 2021 AACR Pancreatic Cancer Special Conference: 2019-2023 ASCO: 2019-2023 ASCO-GI: 2021-2023 DDW: 2019-2023 ESMO: 2019-2024 ESMO World Congress on Gastrointestinal Cancer: 2019-2021, 2023 ESMO Japan: Unavailable in Embase ESMO Asia: 2019, 2020, 2022, 2023 ISPOR EU: 2019-2023 ISPOR US: 2019-2023 ISPOR Asia Pacific: 2020

Abbreviations: AACR, American Association for Cancer Research; ASCO, American Society of Clinical Oncology; ASCO-GI, American Society of Clinical Oncology Gastrointestinal Cancers Symposium; DDW, Digestive Disease Week; ESMO, European Society for Medical Oncology; ISPOR, The Professional Society for Health Economics and Outcomes Research.

Note: Conferences not indexed in Embase were not searched separately, as NK AI does not have this capability. The inclusion of conference abstracts via hand searching was not considered when comparing results between AI-assisted and in-house SLRs.

Table S3. PICOS Eligibility Criteria

Criteria	Include	Exclude
Population	Patients aged ≥ 18 years with metastatic pancreatic adenocarcinoma	• Diseases other than pancreatic adenocarcinoma • Resectable or locally advanced, nonmetastatic disease • Patients receiving maintenance therapy (ie, those with response or stable disease after 4-6 mo) • Pediatric population • Nonhuman population (in vitro, ex vivo, or animal studies)
Intervention/comparators	Any approved or investigational therapy, placebo, standard of care, or no treatment	NA
Outcomes	• Total costs (direct + indirect) • Direct costs • Indirect costs (ie, productivity loss) • Healthcare resource use (eg, hospitalization, outpatient visits, pharmacy use, procedures)	• Outcomes other than those listed (eg, epidemiology, pharmacokinetics, economic evaluations) • Outcomes not separable for the population of interest
Study design	• Observational studies • Cost-of-illness studies • SLRs and narrative reviews	• Clinical trials • Guidelines
Publication type	• Articles, including errata and retractions • Conference abstracts	• Trial protocols • Editorials, comments, notes, and letters
Language	English language	Full-text articles not published in English
Publication date	• Full-text articles: 2014-2024 • Conference abstracts: 2019-2024	• Full-text articles published prior to 2014 • Conference proceedings published prior to 2019

Abbreviations: NA, not applicable; PICOS, population, interventions, comparisons, outcomes, and study design; SLR, systematic literature review.

Table S4. AI Configuration

Characteristics of the AI Screening Model (Robot Screener) ^a	Characteristics of Smart Tag Recommendations ^a
• Gradient-boosted decision tree ensemble. Its hyperparameters, particularly around model complexity (number of trees, tree depth) are optimized using a cross-validation grid search. • The model produces posterior probabilities and is optimized on logistic loss. • Synthetic Minority Oversampling Technique (SMOTE) oversampling is employed as a correction to highly imbalanced classes frequently seen in screening. • The Robot Screener is a machine-learning model that is trained on the records' bibliographic data (ie, time since publication, page count, key words and descriptors), abstract content (ie, n-gram language model, OpenAI text embedding such as text-embedding-(ada-002) to measure how text strings are related), and citation counts (ie, number of citing publications, number of supporting citation statements, number of contrasting citation statements). • If any of these fields are missing from an individual record, it is imputed as if the record is approximately typical to other records in the nest.	• Smart Tag Recommendations is a feature specific to NK. • Smart tags is a content extraction large language model (LLM) that uses Generative Pre-trained Transformer 4 (GPT-4). • This feature identifies the most relevant text in the PDF based on human-generated prompts and returns plain text via a type of search called "fuzzy query," which searches for the most accurate answer to the prompt by measuring edit distance (also known as Levenshtein distance) in the text.

^aFurther details can be found at: https://wiki.nested-knowledge.com/doku.php?id=wiki:support:ai_disclosure

Table S5. Prompts Used for Full-Text and Data Extraction^a

PICOS Criteria	Prompt	Aliases Used^b
Population	Study population	What population was studied? –
	Disease	Did this study include patients with pancreatic cancer? pancreatic ductal adenocarcinoma pancreatic adenocarcinoma ductal carcinoma of pancreas carcinoma of pancreas ductal carcinoma of the pancreas pancreas ADK squamous cell carcinoma of the pancreas adenosquamous carcinoma of the pancreas colloid carcinoma of the pancreas
	Disease stage	Did patients in this study have metastatic cancer? advanced metastatic metastases
Intervention/ comparators	Treatment	What treatment/s were administered to patients? –
	Previous treatment(s)	Did patients receive any treatments prior to enrollment in the study? –
	Line of therapy	What was the line of therapy patients received within the study? –
Outcomes	Outpatient costs	What were the costs associated with outpatient visits? primary care physician doctor laboratory technician specialist
	Indirect costs	What were the data associated with indirect costs of care? sick days absenteeism presenteeism work productivity missed work missed school travel caregiver
	Length of stay	What was the duration of time in which patients had a hospital stay? length of stay duration of hospitalization LOS hospital stay
	Emergency (%)	What is the proportion of patients who had an emergency department or urgent care visit? emergency room emergency department urgent care
	Emergency (No.)	What was the mean or median number and standard deviation or range of emergency department or urgent care visits? emergency room emergency department urgent care
	Outpatient (%)	What is the proportion of patients with outpatient visits (overall or by individual type of outpatient visit (physician, doctor, primary care, specialist, laboratory, etc))? primary care physician doctor laboratory technician specialist
	Outpatient (No.)	What was the mean or median number and standard deviation or range of outpatient visits (overall or by individual type of outpatient visit (physician, doctor, primary care, specialist, laboratory, etc))? primary care physician doctor laboratory technician specialist

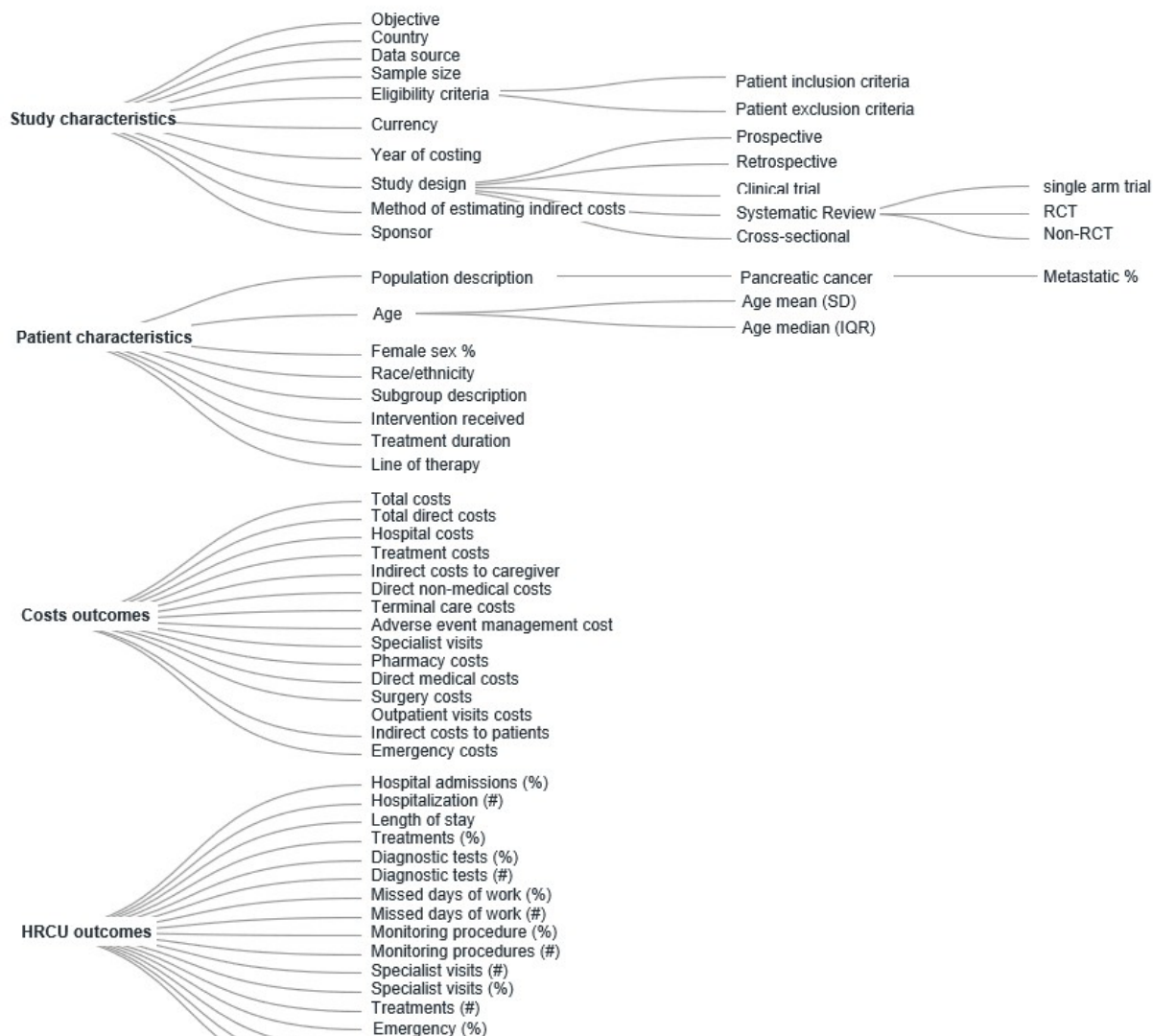
Table S5. Prompts Used for Full-Text and Data Extraction^a

PICOS Criteria	Prompt	Aliases Used ^b
Study design	Study design–general	What is the study design? –
	Study design–prospective	Is this a prospective observational study? –
	Study design–retrospective	Is this a retrospective observational study? –
	Study design–clinical trial	Is this a clinical trial? –
Clinical trial	Single arm trial	Is this a clinical trial with only one treatment group? –
	RCT	Were patients randomized to their treatment groups? –
	Non-RCT	Was this a clinical trial with multiple treatment arms but without randomization of patients to treatment groups? –

Abbreviations: ADK, ***; LOS, length of stay; RCT, randomized controlled trial.

^aA representative selection of prompts is provided. The full library of prompts is available from the authors upon reasonable request.

^bAn alias is an alternative name for the same concept to aid the model.

Figure S2. Hierarchy of Prompts Designed for AI Data Extraction^a

^aA representative selection is provided. The full list of prompts is available from the authors upon reasonable request.