ISPOR EUROPE 2026 · VIENNA · 8–11 NOVEMBER 2026

HTA374

# Evaluating AI-Assisted Time-Aware Evidence Mapping for Health Technology Assessment

## A Pilot Reviewer Audit of Traceability, Temporal Validity, and Jurisdictional Alignment

**Annabelle I. Day[1], Yi-Ting Jonathan Lin[2], Chun-Wei Hsu[1], Kuan-Ju Wang[3], Chao-Hsin Michael Ding[4], Wei-ming Huang[5], Ching-Po Lin[1]**

[1]Institute of Neuroscience, National Yang Ming Chiao Tung University, Taipei, Taiwan; [2]Ph.D. Program of Interdisciplinary Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan; [3]PanCAD.ai Co., Ltd., Taipei, Taiwan; [4]Department of Education and Research, Taipei City Hospital, Taipei, Taiwan; [5]Institute of Health and Welfare Policy, National Yang Ming Chiao Tung University, Taipei, Taiwan

國立陽明交通大學
NATIONAL YANG MING CHIAO TUNG UNIVERSITY

## Towards Traceable, Time-Aware, and Reviewable HTA Evidence

**91.7%** Source traceability

**86.8%** Temporal validity

A pilot reviewer audit across 8 oncology medicines and 288 structured evidence fields.

## 1 Objectives

- **Artificial intelligence (AI)** is increasingly being explored to support HTA evidence identification, synthesis, and report preparation.
- AI-generated HTA evidence requires **verifiable sources**, **temporal validity**, and **jurisdictional alignment**.
- This pilot evaluated **claim-level reviewability** and the remaining **reviewer correction burden**.

## 2 Methods

Workflow from source documents to independent reviewer audit

**1 Source documents**

8 oncology medicines: regulatory documents, pivotal trials and HTA reports.

**2 Evidence extraction and five-tag mapping**

Sources split into single claims, each tagged with five metadata fields (Box 1). Dated claims are linked in a temporal evidence graph.

BOX 1 · ANATOMY OF ONE AUDITABLE CLAIM

**“In the FLAURA trial, osimertinib was compared with gefitinib or erlotinib.”**

| | |
|---|---|
| Source type | **HTA report (CDE)** |
| Document date | **2024-08** |
| Page or location | **p. 2-3** |
| Evidence domain | **Comparators** |
| Jurisdiction | **Taiwan (NHI)** |

**3 288 structured evidence fields**

Indication, population, comparators, trials, endpoints, safety, regulatory status, HTA recommendations, reimbursement restrictions and uncertainty.

**4 Independent reviewer audit**

2 trained reviewers checked every field on three criteria:

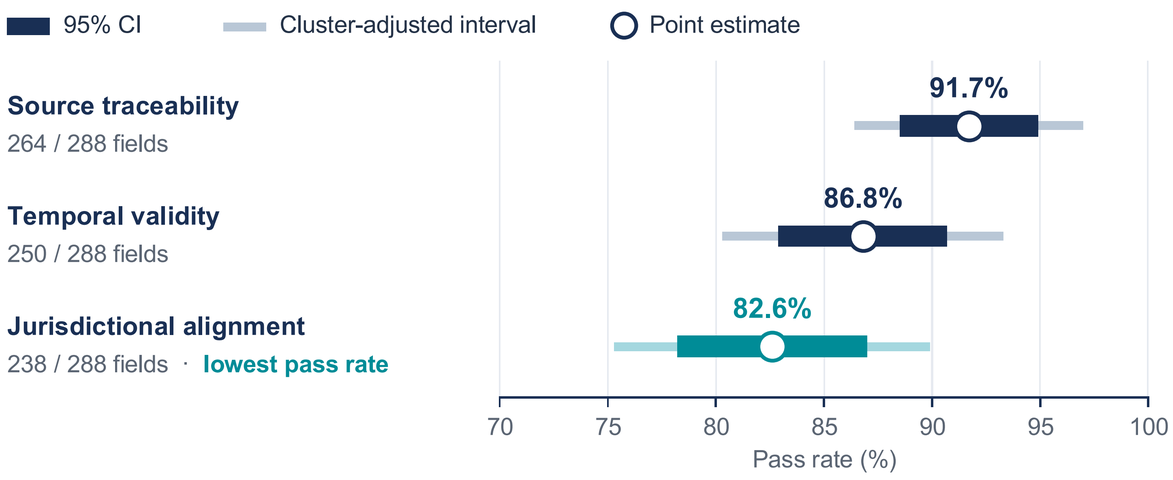
| | |
|---|---|
| **Source traceability** | Can the cited source be located? |
| **Temporal validity** | Valid at the index date? |
| **Jurisdictional alignment** | Applicable to the target jurisdiction? |

**5 Correction grading and cluster-aware analysis**

Corrections graded as none, minor or major; confidence intervals account for clustering within medicines.

## 3 Results

**Figure 1.** Primary outcomes, pass rate of 288 fields




Pale intervals adjust for clustering within medicines. Counts rounded from percentages.

Jurisdictional alignment had the lowest pass rate; mismatches involved comparators, reimbursement restrictions and outdated recommendation wording.

**Reviewer correction burden.** Share of 288 fields

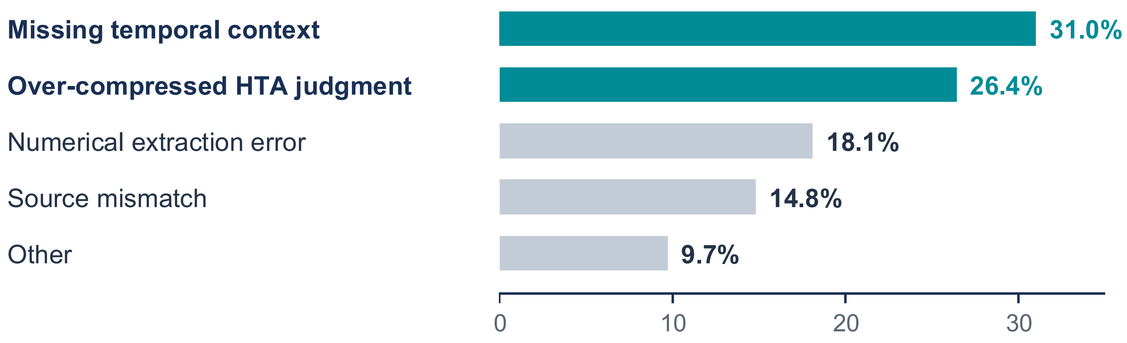
**29.8%** **of fields required reviewer correction**




**Figure 2.** Error types, share of recorded errors

**57.4%** **Top two error types combined**

Missing temporal context + over-compressed HTA judgment




Recorded errors (%). Error denominator and category overlap require confirmation.

AN ERROR THAT ACCURACY CANNOT CATCH

**Source:** “Recommended, with conditions, in a defined subgroup.”

**As mapped:** “Recommended.”

Factual accuracy: passes. **HTA review: fails.**

**Reviewer agreement**

**κ = 0.78**

The high traceability pass rate may affect κ. Raw agreement, PABAK and per-judgment results are referenced in the conference handout.

**Generation time**

**11.4 min** per medicine

Median generation time excludes reviewer audit. No manual comparator arm: feasibility, not comparative efficiency.

## 4 Conclusions

**01**

**Time-aware evidence mapping** supports transparent, claim-level HTA review.

**02**

Residual errors, especially jurisdictional and temporal issues, require **human verification**.

**03**

Future validation should extend across **therapeutic areas and jurisdictions**.

### Research Collaboration & Future Validation

**Cross-jurisdictional evaluation · Broader therapeutic validation · Collaborative HTA research**

National Yang Ming Chiao Tung University, Taipei, Taiwan · **Seeking external validation partners**

**Study materials & research contact**

Scan the QR code

**Abbreviations:** CI, confidence interval; HTA, health technology assessment; PABAK, prevalence-adjusted bias-adjusted kappa.