# Delivered in an Hour, But at What Cost? The Silent Costs of AI in HEOR and the Case for Outcome-Based Pricing

MSR236

**Tushar Srivastava**[1], Shilpi Swami[1], Kunal Hriday[2]

[1]ConnectHEOR, London, UK[1] ; [2]ConnectHEOR, Delhi, India.

Email: tushar.srivastava@connectheor.com

connectHEOR

> **Across 10 completed AI-assisted deliverables in four work types, counting only analyst hours overstated the benefit once token cost and added senior review were included; net economics were favourable for repeatable, specified work and weak for one-off judgement-heavy work, which argues for pricing the validated deliverable rather than hours.**

## 1 BACKGROUND

Claims that **AI saves time** in health economics and outcomes research (HEOR) are now routine. Large surveys of legal, tax and audit professionals estimate hours freed per practitioner and convert them directly into annual value per head.[1]

- Consultants using a frontier model complete more tasks, faster and at higher quality inside its capability frontier, but did materially worse outside it.[2]
- Over-reliance emerged exactly where closer supervision was needed, which is the part the productivity arithmetic tends to drop.

**GAP:** two silent costs are missing from the headline. **Token cost** is often ignored because unit prices fell by more than two orders of magnitude in about two years.[3] **Senior verification** is required, as health technology assessment (HTA) guidance makes human oversight a standing requirement.[4–6]

## 2 OBJECTIVES

The study set out to quantify the net economics of AI-assisted HEOR once both silent costs are counted, and then to interpret what the resulting cost structure implies for how HEOR consulting prices its work.

**1 QUANTIFY**

Measure effort and cost with and without AI, then subtract token spend and added senior review from the gross analyst time saved.

**2 INTERPRET**

Establish whether the hour-based billing on which HEOR consulting still runs reflects the value of AI-assisted work.

## 3 METHODS

The analysis was conducted across completed AI-assisted deliverables, so the unit of analysis was work that had actually been produced and signed off rather than a simulated or benchmarked task. The deliverables spanned four distinct work types: systematic literature review, cost-effectiveness modelling, model validation, and technical reporting.

**STEP 1** Gross analyst-time saved **–** **STEP 2** Token spend **–** **STEP 2** Added senior review and sign-off **=** **STEP 3** Net cost

- Effort and cost were measured **with and without AI** on the same deliverables, giving an internal comparison rather than a remembered baseline.
- Step 1 recorded gross analyst-time saved, which is the quantity conventional time-savings claims report and then stop at.
- Step 2 subtracted token spend and the added senior review, validation and sign-off time, and step 3 gave net cost.
- Time-based billing was analysed in its own right, testing whether true cost and billed time diverge in a patterned way.

## 4 RESULTS

Counting only analyst hours **overstated the benefit in every deliverable**, and in no case did the conventional and net calculations agree.

| | |
|---|---|
| **0** | cases in which the hours-only and net calculations agreed |
| **4 of 4** | work types in which counting analyst hours alone overstated the benefit |
| **2** | cost regimes: repeatable, specified work and one-off, judgement-heavy work |

### Where the effort went

| Junior production | Senior review | Token cost |
|---|---|---|
| compressed sharply | more time than before | non-trivial at scale |

AI compressed junior production time sharply, but a senior health economist spent **more time reviewing AI output** than before. The hour removed is a junior hour and the hour added is a senior one, so the overstatement is systematic.

### Net economics by task type

| Dimension | Repeatable, high-volume, specified | One-off, judgement-heavy |
|---|---|---|
| **Token and review cost** | Amortised across many items | Nothing to amortise across |
| **Senior verification** | A fraction per additional item | Hours, even for a deliverable produced in an hour |
| **Net position** | **Strongly favourable, as the saving compounds** | **True cost and billed time diverged sharply** |

Once a check or reporting structure is specified and validated, each new item carries compute cost but little verification, so **the saving compounds**. One-off work has nothing to amortise across, so true cost and billed time came apart.

## 5 DISCUSSION

When production is compressed and verification is not, **senior judgement** becomes the expensive input, and it scales with the novelty of a task rather than its volume.

- Amortisation works at the repeatable end and is unavailable at the bespoke end, so the same tooling produces opposite economics.
- Hour-based billing prices the input AI has made cheap, undercharging for embedded expertise and for the value the client buys.
- HTA guidance makes evidenced human oversight a fixed obligation, consistent with AI that is assistive and governed.[4,5,7]
- Limitations: results reflect one firm's seniority mix, rates and time recording, token prices keep moving, the pricing conclusion is untested in the market, and the authors sell HEOR services, so the pricing argument carries a potential interest.

AI, artificial intelligence; HEOR, health economics and outcomes research; HTA, health technology assessment; NICE, National Institute for Health and Care Excellence.

## 6 CONCLUSION

When a task is **delivered in an hour** but consumes senior review and real compute, **hour-based billing misprices** the work. The benefit of AI is **real but conditional**, concentrated in repeatable work where verification amortises.

- **Repeatable work:** verification amortises and the saving compounds.
- **Bespoke work:** senior review and tokens offset the time saved.
- **AI ROI claims** must count tokens and senior review, not hours alone.

**Implication: move from hours towards outcome- and value-based pricing, charging for the validated deliverable and the decision it supports.**

REFERENCES

1. Thomson Reuters Institute. Future of Professionals Report. 2025.
2. Dell'Acqua F, et al. Organization Science. 2026;37.
3. Stanford HAI. Artificial Intelligence Index Report. 2025.
4. NICE. Use of AI in evidence generation: position statement. 2024.
5. Fleurence RL, et al. Value Health. 2025;28(2).
6. Fleurence RL, et al. Value Health. 2025;28(11).
7. Srivastava T, et al. Adv Ther. 2025;42.

**Financial disclosure:** Authors are employees of ConnectHEOR Limited. No external funding received. No conflict of interest to declare.

**Poster presented at ISPOR Europe 2026 in Vienna, Austria, 8-11 November 2026**