EE592

Performance Evaluation of Explainable Artificial Intelligence to Replicate a Systematic Review and Meta-Analysis of Cost-Effectiveness Analysis Findings

Jiafan Chen, MS[1], Rahul Mudumba, MHS[1], Joshua Morriss, PhD[2], William V. Padula, PhD[1,3]

[1]Department of Pharmaceutical and Health Economics, Alfred E. Mann School of Pharmacy, University of Southern California, Los Angeles, CA, USA

[2]Ziplitics, Inc., Midlothian, VA USA

[3]Leonard D. Schaeffer Center for Health Policy and Economics, Los Angeles, CA, USA

BACKGROUND & OBJECTIVE

We aimed to measure the cost-effectiveness of vaccines for infectious diseases in the U.S. While vaccines are generally cost-effective, their economic value from U.S. healthcare sector and societal perspectives remains under-evaluated[1]. Our objective was to use and evaluate the performance of the Literature Review Network (LRN), an explainable AI (XAI), to automate a health outcomes systematic literature review (SLR) and meta-analysis on this topic.

METHODS

Search Strategy:
This SLR was conducted via PubMed using open-access, U.S.-based cost-effectiveness analyses; vaccines were the primary intervention, from a societal/healthcare sector perspective.

	Search Strategy
Inclusion Criteria	(((vaccine) OR (vaccines) OR (vaccination) OR (immunization))) AND (((cost-effectiveness) OR (cost effectiveness) OR (cost-utility) OR (cost utility) OR (cost-benefit) OR (cost benefit) OR (economic evaluation) OR (economic assessment) OR (health technology assessment) OR (hta))) AND (((icer) OR (icur) OR (icers) OR (incremental cost effectiveness ratio) OR (incremental cost utility ratio))) AND (((qaly) OR (qalys) OR (daly) OR (dalys) OR (qale) OR (qualy) OR (quality adjusted life year) OR (quality-adjusted life year) OR (quality adjusted life expectancy) OR (quality-adjusted life expectancy) OR (disability-adjusted life year))) AND (2013/01/01:2023/09/18[dp])
Exclusion Criteria	(((vaccine) OR (vaccines) OR (vaccination) OR (immunization))) AND (((cost-effectiveness) OR (cost effectiveness) OR (cost-utility) OR (cost utility) OR (cost-benefit) OR (cost benefit) OR (economic evaluation) OR (economic assessment) OR (health technology assessment) OR (hta))) AND (((icer) OR (icur) OR (icers) OR (incremental cost effectiveness ratio) OR (incremental cost utility ratio))) AND (((qaly) OR (qalys) OR (daly) OR (dalys) OR (qale) OR (qualy) OR (quality adjusted life year) OR (quality-adjusted life year) OR (quality adjusted life expectancy) OR (quality-adjusted life expectancy) OR (disability-adjusted life year))) AND (2013/01/01:2023/09/18[dp]) AND (((non-US) OR (education program) OR (administration) OR (non-English)))

Table 1. Search strategy used by LRN. Excluded studies served as a "ground-truth" negative data set.

Explainable Artificial Intelligence (XAI):
LRN (v1.5) classified studies using proprietary generative-discriminative models, a word embedding model incorporating the UMLS Metathesaurus, and a meta-heuristic wrapper[2]. Leveraging a large language model (LLM) based on GPT-4-turbo and retrieval-augmented generation (RAG), LRN-included reports were utilized for SLR writing[3].

XAI-led Systematic Literature Review:
LRN (v1.5) underwent training via reinforcement learning with human feedback (RLHF) over 3 iterations. A final LRN model was deployed to label the entire corpus. During an iteration, the model incorporated feedback from 21 abstracts screened by 2 human researchers, as either INCLUDE or EXCLUDE, and adjustments to the rules. Explainability metrics described concepts LRN used in its decision-making processes. Final included studies were determined by the researchers; a senior health economist broke ties.

Rule #	Rule	Label	Iteration Added
1	EU	EXCLUDE	1
2	non-English	EXCLUDE	1
3	US	INCLUDE	1
4	cost-effectiveness studies	INCLUDE	1
5	animal	EXCLUDE	1
6	vaccine	INCLUDE	1
7	education	EXCLUDE	1
8	covid	INCLUDE	1
9	africa	EXCLUDE	2
10	asia	EXCLUDE	2
11	europe	EXCLUDE	2
12	cancer	EXCLUDE	2
13	infection	INCLUDE	2
14	qaly	INCLUDE	2
15	daly	INCLUDE	2
16	icer	INCLUDE	2
17	immunization	INCLUDE	2
18	economic	INCLUDE	2

Table 2. User-defined parameters across RLHF iterations. "Iteration added" = decisions by researchers during screening.

XAI-led Meta-Analysis:
Post-SLR, LRN (v2.0) was employed to perform a meta-analysis. Using a fixed-effects inverse-variance weighted model with Fieller's theorem for 95% confidence intervals[4], the analysis assessed the cost-effectiveness of vaccines through ICERs, with a WTP threshold of $150,000 (USD/QALY). Data were extracted via LRN (v2.0), with structured tables supported by an LLM based on GPT-4-turbo.

RESULTS

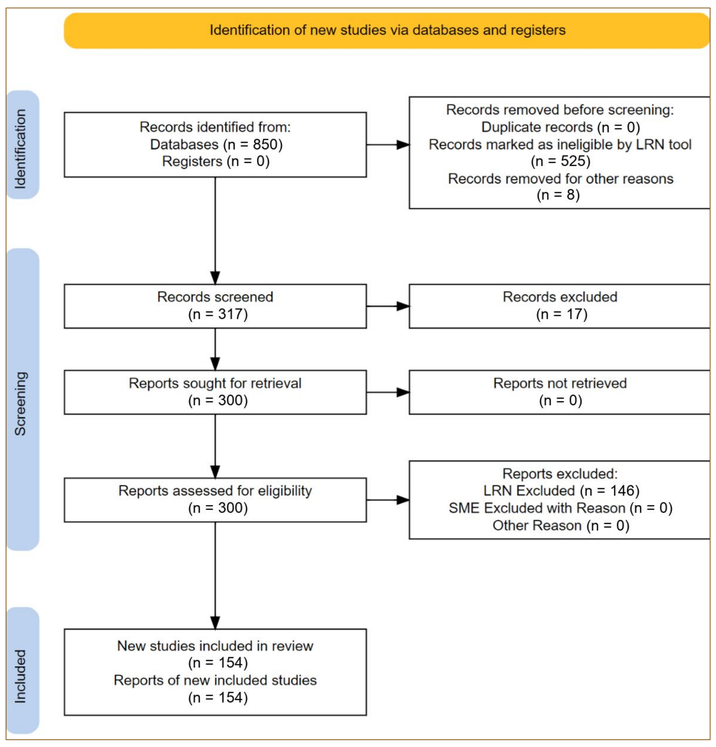

Figure 1. Auto-populated PRISMA 2020 flow diagram. After 3 iterations, researchers excluded 17 records and LRN excluded 146 full-text reports. Studies deemed ineligible by LRN (n = 525) were those that met the exclusion criteria; records removed for other reasons (n = 8) were linguistically incompatible with the NLP processing capabilities of LRN v1.5 (studies written in Chinese or Russian).

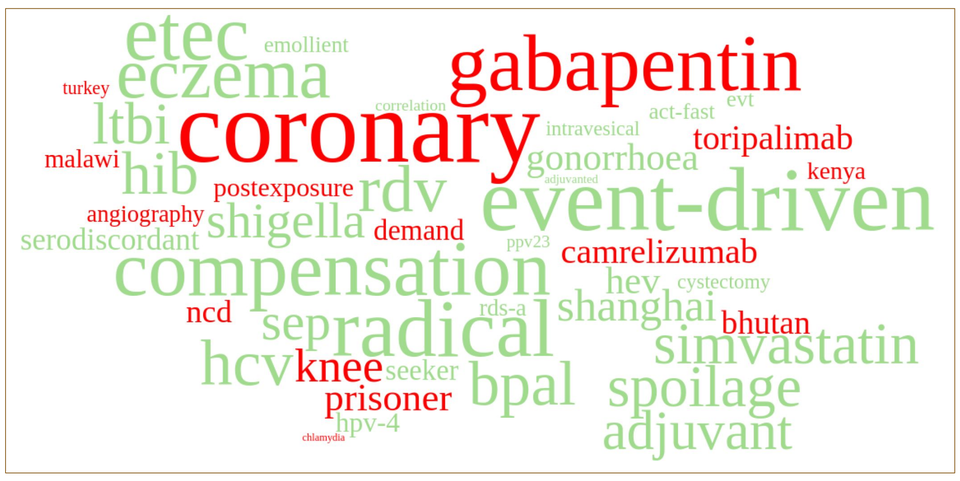

Figure 2. Explainability word cloud and key vaccine economic insights utilized by LRN. Important parameters determined by LRN in its decision making processes. Concept size correlates with its frequency; color indicates relevance to classification: green (INCLUDE) and red (EXCLUDE). From 3rd iteration of RLHF.

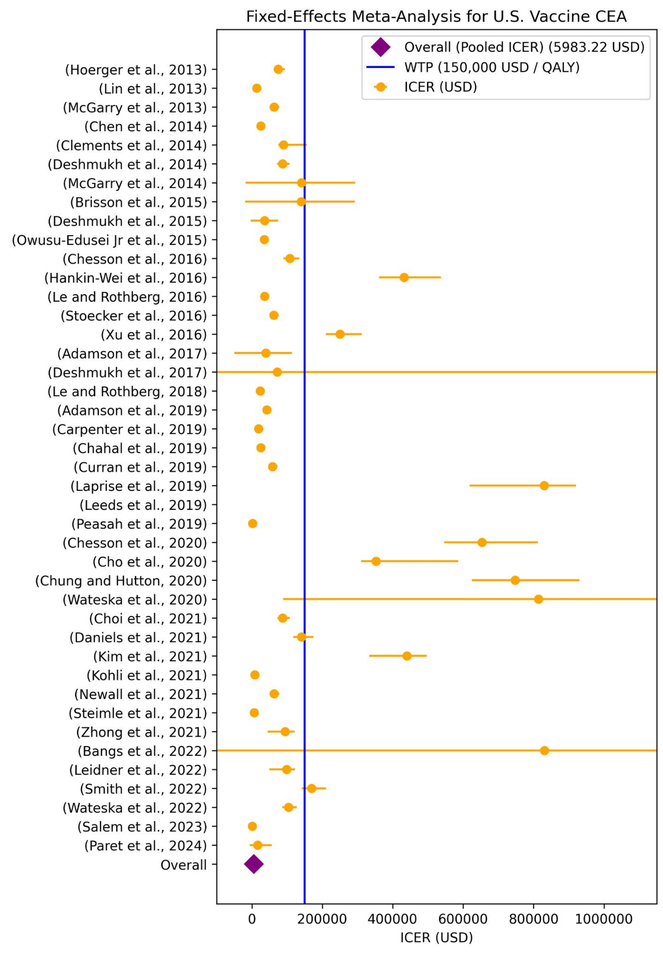

Figure 3. Vaccine cost-effectiveness from the US healthcare sector perspective.

Task	Date of Human Work Session	Human Labor Time (min)	(Date of Model Start : End)	Runtime Start (hr:min:sec)	Runtime End (hr:min:sec)	Total Runtime (hr:min:sec)
Configuration	2023 / 09 / 18	120	2023 / 09 / 18 : 2023 / 09 / 20	16:44:43	13:39:27	44:54:44
RLHF Iteration 1	2023 / 09 / 20	60	2023 / 09 / 20 : 2023 / 09 / 27	20:21:45	14:33:13	162:11:28
RLHF Iteration 2	2023 / 09 / 27	60	2023 / 09 / 27 : 2023 / 09 / 29	20:29:15	18:55:01	46:25:46
RLHF Iteration 3	2023 / 10 / 04	60	2023 / 10 / 09 : 2023 / 10 /11	16:38:28	16:35:42	47:57:14
Deployed Model	-	-	2024 / 04 / 15 : 2024 / 04 / 15	16:51:57	17:26:20	00:34:23
Total Human Labor Time (mins):		300	Total LRN Computation Time (mins):			18123.58

Table 3. Productivity metrics for human labor versus computational time via LRN systematic literature review. All time reported in Coordinated Universal Time (UTC), reported in 24-hour time.

RESULTS

Class	Recall	Precision	F-score
Iteration 1			
INCLUDE	50.00%	56.67%	53.13%
EXCLUDE	41.38%	48.00%	44.44%
Iteration 2			
INCLUDE	89.19%	63.46%	74.16%
EXCLUDE	17.39%	50.00%	25.81%
Iteration 3			
INCLUDE	73.53%	64.10%	68.49%
EXCLUDE	46.15%	57.14%	51.06%

Table 4. Performance metrics for XAI classifications. RLHF iteration 3, or the "highest performing iteration" achieved a Cohen's kappa of 0.2014; highest INCLUDE label performance was in iteration 2.

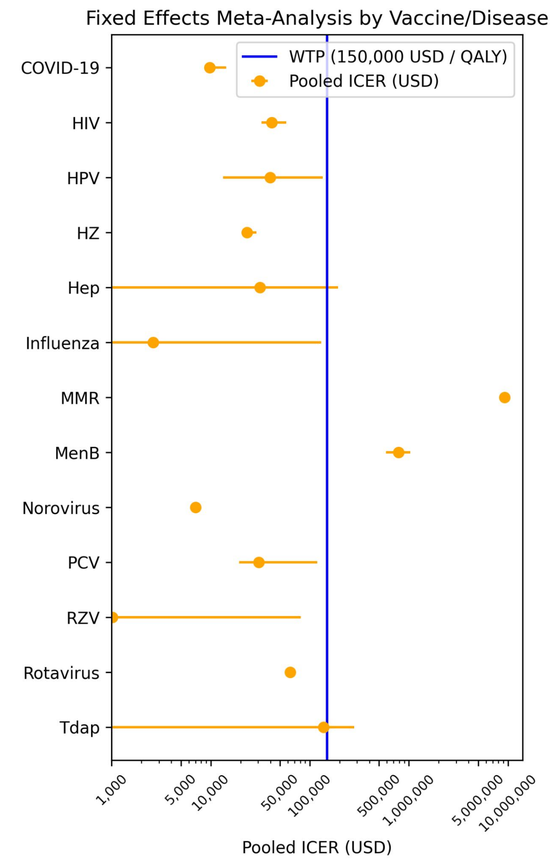

Figure 4. XAI-led meta-analysis reveals vaccines cost-effective by disease state from a healthcare sector perspective.

Vaccine / Disease	Pooled ICER	Standard Error	95% CI (Fieller's Theorem)
COVID-19	9779.93	7868.53	[9302.02, 14292.31]
HIV	41305.26	29011.68	[32687.82, 57806.99]
HPV	39664.65	2572.64	[66014.19, 135670.38]
HZ	23249.77	14858.36	[25137.30, 28866.41]
Hep	31270.06	22827.43	[132656.43, 192223.08]
Influenza	2632.34	1852.07	[61605.98, 129755.89]
MMR	9200000	5798750	[-, -]
MenB	785926.19	746861.65	[585636.99, 1034769.36]
Norovirus	7028	7124	[-, -]
PCV	30545.78	11753.63	[41825.41, 119072.72]
RZV	1018.42	976.038	[40340.43, 81104.58]
Rotavirus	63228	63227.95	[-, -]
Tdap	137420.62	13744.068	[-17223.00, -5207.32]

Table 5. Pooled ICERs across CEAs by disease state. HPV = human papillomavirus, HZ = herpes zoster, Hep = hepatitis a/b, MMR = measles, mumps, and rubella, MenB = meningococcal serogroup B, PCV = pneumococcal conjugate vaccine, RZV = recombinant zoster vaccine, Tdap = tetanus, diphtheria, and pertussis.

CONCLUSIONS

Our XAI-led SLR and meta-analysis found vaccines to be highly cost-effective from a U.S. healthcare sector perspective, particularly for high-burden diseases such as COVID-19, HIV and HPV at standard WTP thresholds; this supports their preferential status on formularies for health and cost benefits. LRN achieved high performance in 240 minutes. In 300 minutes, the final LRN model screened 850 studies, selecting 154; 77 confirmed by human reviewers, and 45 studies retrieved for final analysis. This study demonstrates that explainable AI, like LRN, is an effective tool for advancing population health.

REFERENCES

1. Leidner AJ, Murthy N, Chesson HW, et al. Cost-effectiveness of adult vaccinations: A systematic review. Vaccine. 2019;37(2):226-234. doi:10.1016/j.vaccine.2018.11.056
2. Bodenreider, O. (2004). The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Research, 32*(Database issue), D267–D270. https://doi.org/10.1093/nar/gkh061
3. OpenAI, Achiam, J., Adler, S., … Zoph, B. (2023). *GPT-4 Technical Report* (arXiv:2303.08774). arXiv. https://doi.org/10.48550/arXiv.2303.08774
4. Polsky D, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. Health Econ. 1997;6(3):243-252. doi:10.1002/(sici)1099-1050(199705)6:3<243::aid-hec269>3.0.co;2-z