WHICH GENERATIVE AI METHOD USED FOR HIGH SPECIFICITY: A METHODOLOGICAL COMPARISON FROM THE SYSTEMATIC LITERATURE REVIEW OF THE BURDEN OF INFLUENZA IN FRANCE

MSR224

LAMARSALLE Ludovic[1], LEMAITRE Magali[2]
[1]Healstra, Lyon, France | [2]Health data Expertise, Genissieux, France |
Contact Information:
Ludovic LAMARSALLE, MSc, PharmD
llamarsalle@healstra.com

MyHA

INTRODUCTION

1- **Systematic literature reviews (SLRs)** are essential for evidence synthesis in health economics and outcomes research, but face increasing challenges due to exponential growth in published literature. Manual screening of thousands of abstracts is time-consuming, resource-intensive, and subject to reviewer fatigue and inconsistency.

2- Generative artificial intelligence (AI) offers promising solutions to enhance SLR efficiency while maintaining rigorous selection standards. However, optimal AI implementation strategies remain unclear: single vs. multiple models, batch vs. individual processing, human vs. algorithmic arbitration.

OBJECTIVES

To evaluate the efficiency and accuracy of two artificial intelligence (AI) methodologies for conducting systematic literature reviews (SLR) on **influenza burden among elderly populations** in France, comparing performance metrics, resource utilization, and consistency of findings.

METHODS - AI Method comparaison for Literature review

Systematic literature review - Pubmed
From january 2010 to march 2025 influenza-related terms, France-specific terms, yielded 2,060 abstracts

Inclusion criteria
Elderly population (aged 60+), sample size over 10,000, conducted in France, and addressing at least one component of influenza burden : epidemiological, clinical, economic, or humanistic

Characteristic	Method 1: GPT-4o & Mistral Large	Method 2: Claude-3.7-Sonnet & GPT-4o
Processing type	Batch processing, 72 cycles (batches of 25)	Sequential processus, individual evaluation : Claude 3.7 Sonnet then GPT-4o
Number of abstracts analyzed	1,810 (250 empty)	1,817 (243 empty)
Time to completion & selection	One hour, GPT-4o: 31, Mistral : 36	Two hours, agreement in 99.2% (disagreement for 15 abstracts)
Arbitration / Final selection	Human, 30 articles retained (37 false positive)	**Mistral** arbitration: 13 articles retained (2 false positive)

RESULTS

Superior precision (method 2)

Model consensus approach showed 56% better specificity with 94.6% reduction in false positives (37→2) and 99.2% inter-model agreement.

Efficiency trade-off

Method 1: Faster (2 sec/abstract) but lower precision.
Method 2: Double time (4 sec/abstract) but 10-fold reduction in validation burden.

CONCLUSION

High-volume validation

Both methods successfully processed ~1,800 abstracts, demonstrating scalability for comprehensive SLRs across 15 years of literature.

Recommendation

Use Method 2 for SLRs requiring high specificity and rigorous selection criteria. Processing time investment (2x) yields substantial reduction in false positives (18.5x).

A disproportionate influenza burden for elderly populations

155-350 M€ per season

20 % rehospitalizations

25,000 - 55,000 hospitalizations

Hospital Deaths: Elderly 80%

Excess hospitalizations: Elderly 70%

Associated Costs: Elderly 77%

References

Demont C, Mouaddin NE, Chillotti L, Bénard S, Salhi A, Uhart M, et al. Fardeau de la grippe, de la COVID-19 et du VRS dans les hôpitaux français: Analyse des données du PMSI de 2018 à 2023. Médecine Mal Infect Form. 1 juin 2025;4(2, Supplement):S127-8.

Lemaitre M, Fouad F, Carrat F, Crépey P, Gaillat J, Gavazzi G, et al. Estimating the burden of influenza-related and associated hospitalizations and deaths in France: An eight-season data study, 2010-2018. Influenza Other Respir Viruses. juill 2022;16(4):717-25.

Paternoster M, Masse S, van der Werf S, Lina B, Levy-Bruhl D, Villechenaud N, et al. Estimation of influenza-attributable burden in primary care from season 2014/2015 to 2018/2019, France. Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol. juin 2021;40(6):1263-9.

Romane Le Goff, Andrea Contini, Pascal Crepey, Jacques Gaillat, Gaetan Gavazzi, Odile Launay, Anne Mosnier, Léa Antoniali, Hélène Bricout. Impact de la Grippe sur le Parcours de Soins des Personnes Âgées en France - Une étude sur les données issues du chainage EMR/SNDS. JNI 2025;

Pivette M, Nicolay N, de Lauzun V, Hubert B. Characteristics of hospitalizations with an influenza diagnosis, France, 2012-2013 to 2016-2017 influenza seasons. Influenza Other Respir Viruses. mai 2020;14(3):340-8.