AI to Fully Automate Systematic Literature Reviews (SLRs) and HTA Dossiers: Is It Viable, Wise, and Valuable?

MSR 71

L.Z. BEATSON[1], **L.C. GOLD**[1], and **M. LOOTUS**[1]

[1]**Tehistark LTD**, London, United Kingdom

www.tehistark.com

SOURCE DATA

1 INTRODUCTION

1.1 Problem and Objective

HTA teams face unprecedented challenges: submission timelines shortening while evidence volumes grow exponentially. With limited expert availability and increasing submission complexity, we address three critical questions:

1. Can AI fully automate HTA dossier creation with acceptable quality?
2. How does automation impact the contribution of human expertise?
3. What unique value can AI bring to evidence synthesis and submission quality?

To answer these questions, we conducted a validation study using a real-world case: the 2019 EUNET HTA of Siponimod for Secondary Progressive Multiple Sclerosis (SPMS) [1].

1.2 AI Capabilities

Our solution leverages state-of-the-art AI Models:

- Advanced LLMs (GPT-4o, Claude 3.5, LLAMA3)
- Specialized planning and extraction models
- Custom-trained dossier synthesis engines

These models are enhanced through:

- Comprehensive prompt engineering
- Rigorous validation frameworks
- Domain-specific training, expert quality controls

1.3 AutoSLR & AutoDossier

An AI-native platform for rapid reviews, TLR-s, SLR-s, and dossiers: faster, more transparent, better.

- End-to-end automation: Plan, Search, Screen, Extract, Write
- LLM audit trail system with data provenance
- Strategise, Route, Write, Revise, Update
- Secure cloud infrastructure
- Template-driven workflows for: rapid reviews, targeted literature reviews, full systematic reviews, regulatory submissions (IND, NDA/MAA/BLA, HTA)

Built-in connections to past studies, templates, and guidelines

2 METHOD

2.1 Study Design

We conducted a three-arm validation study using the 2019 EUNET HTA submission [1] for Siponimod in SPMS as reference case:

1. **Full Automation**: AI plans and executes the entire process of dossier planning and writing, with no human input.
2. **Hybrid Approach**: Human planning with AI execution
3. **Traditional Process**: 2019 EUNET HTA as reference standard

2.3 Evaluation Methods

- Protocol quality assessment
- Search strategy comprehensiveness
- Screening accuracy (recall/precision)
- Extraction accuracy against ground truth
- Final dossier quality evaluation

Table 1: Automation Details

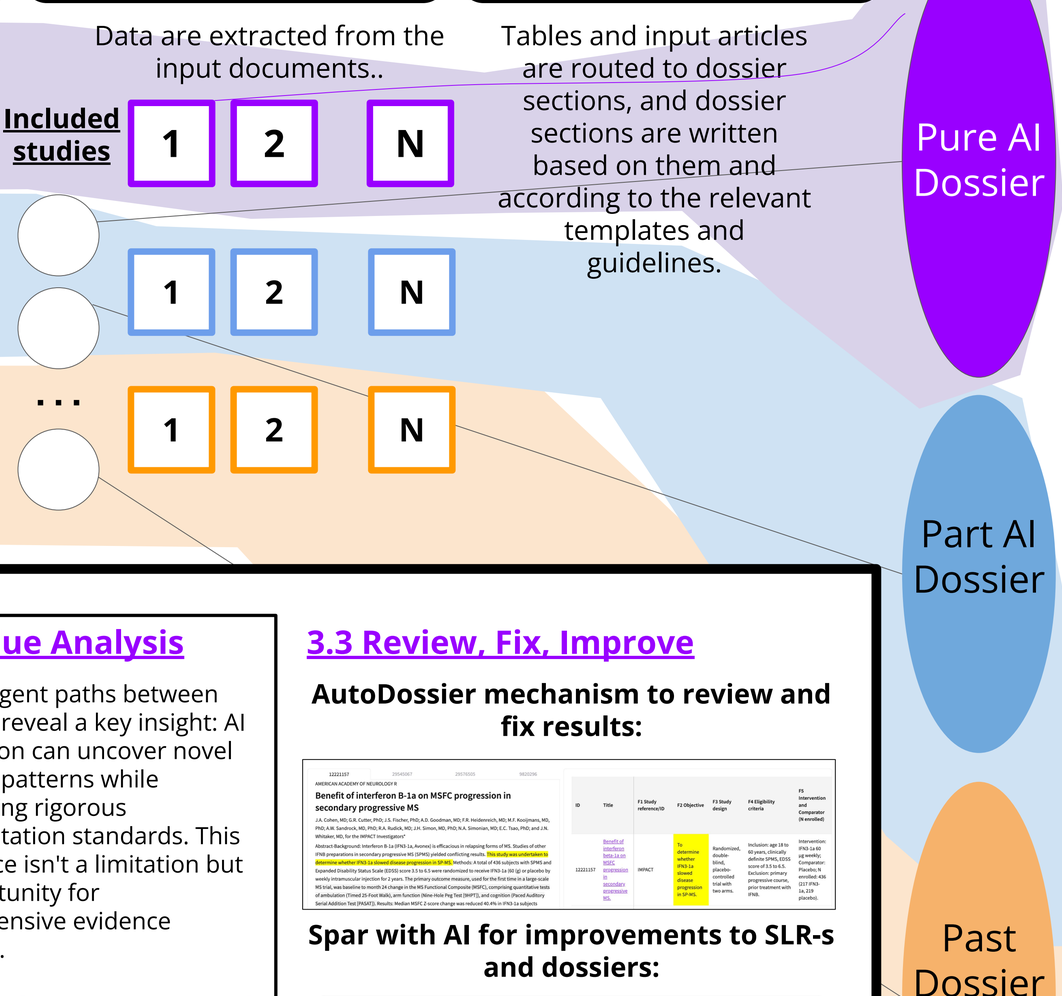
	Step 1: Protocol Development	**Step 2**: Literature Search	**Step 3**: Screen Articles	**Step 4**: Extract	**Step 5**: Write Dossier Sections
Full-Auto: Pure AI.	AI generated HTA protocol and SLR protocol (search plan, screening criteria, and extraction plan)	AI-generated query and automated article search	Automated TA screening + automated FT screening with AI protocol	Automated extraction to AI-designed tables.	Fully automated writing, according to HTA template and guidelines.
Part-Auto: Humans with AI.	Human protocol with AI support	Human query with AI refinement	Automated TA screening + automated FT screening with human protocol	AI extraction to human-designed tables	Fully automated writing, according to HTA template and guidelines.
Manual Humans.	Human-written	Human query with manual searches	Manual screening with human protocol	Manual extraction to human- designed tables.	Human-written

Step 1: Protocol Development | **Step 2:** Literature Search | **Step 3:** Screen Articles | **Step 4**: Extract | **Step 5**: Write Dossier Sections

Arm 1: **Full-Auto**. 1% the time. 100x the volume.

Arm 2: Part-Auto, Part Humans.

Arm 3: **Traditional**. Published HTA from 2019 on MS.



3 RESULTS

Table 2: Automation Details

	Step 1: Protocol Development	**Step 2**: Search articles	**Step 3**: Screen articles	**Step 4**: Extract to tables	**Efficiency Metrics**
Full-Auto	INCLUSION • Population: Studies must include patients diagnosed with secondary progressive multiple sclerosis (**SPMS**). • Intervention: Studies must investigate the effects of Siponimod as a treatment. • Outcome: Studies must report on clinical efficacy and/or safety outcomes of Siponimod treatment. EXCLUSION • Non-human studies: Exclude studies that are not conducted on human subjects. • Non-SPMS population: Exclude studies that do not focus on patients with secondary progressive multiple sclerosis. • Other Interventions: Exclude studies that do not focus specifically on Siponimod as the intervention.	**The query**: 75-term query fully autonomously written by AI, optimised research question and protocol: 18 P-terms, 15 I-terms, 11 C-terms, 17 O-terms, 14 S-terms. **Sources:** PubMed **Records returned: 85**	Records identified: **85** Input to TA: 85 Included by TA: 31 FT-s auto-obtained: 13 FT-included: **12 (disjoint with manual)** Out of the included articles, 1 is a protocol (the inclusion exclusion criteria did not ask these to be exclude, 1 is in German - but the extraction still worked.	Data Extraction Accuracy (Full Auto): • 60% perfect match • 33% partial match • 7% requiring review • Zero hallucinations	85 records → 12 studies 5 minutes
Part-Auto	INCLUSION • Adult: The study population includes adults aged 18 years and older. • SPMS: Patients are diagnosed with secondary progressive multiple sclerosis (**SPMS**). • Intervention: The study must investigate at least one of the following interventions: *(list of 25 treatments including Siponimod)* • Comparator: *(conditions on placebo, in 50 words)* • Outcomes: The study must report any efficacy, health-related quality of life (HRQoL) or safety outcomes, including: *(plus 115 words more)* • Study Design: The study design one of the following allowed types: (+70) • English: The abstract or full text is in English. EXCLUSION • Mixed Population: The study reports eligible outcomes in a mixed population, without separately reporting data for the population of interest (unless more than 80% of study population are adults with SPMS) • Non-human: The study has non-human subjects.	**The query**: 69 part composite query written by the original study team: 6 I&P-terms, 32 S-terms, 30 I&C-terms. **Sources**: PubMed **Records returned**: 2715	Records identified: **2715** Input to TA: 336 TA-included: 36 FT-s obtained: 25 FT-included: **17** **TA Recall = 0.92** **FT recall = 0.90**	Hybrid Performance: • 58% perfect match • 35% partial match • 8% poor match • Zero hallucinations	2,715 records → 13 studies 1 week
Manual	**The query**: Same as last row. **Sources**: PubMed + other **Records returned**: 3478 total (2726 from PubMed)	Records identified: **3478** Input to TA: 3212 TA-included: 341 FT-s obtained: 341 FT-included: 97 (23 studies)	The AI discovered (1) page 55 NCT number wrong for ASCEND trial, (2) two cells that were missing from the original dossier draft IMPACT study on page 58.		3,478 records → 23 studies 8 weeks

3.1 Value Analysis

The divergent paths between methods reveal a key insight: AI automation can uncover novel evidence patterns while maintaining rigorous documentation standards. This divergence isn't a limitation but an opportunity for comprehensive evidence synthesis.

3.2 Quality Framework

Every output undergoes:

- Automated consistency checking
- Cross-reference validation
- Source traceability verification
- Expert review capability
- Rapid iteration cycles

3.3 Review, Fix, Improve

AutoDossier mechanism to review and fix results:

Spar with AI for improvements to SLR-s and dossiers:

1: "Missing comparison with Phase III data from EXPAND trial. Consider adding direct efficacy comparison with primary endpoints."

2: "Safety data presentation could be strengthened by including subgroup analyses, especially for elderly patients and those with comorbidities."

3: "Statistical methodology section needs more detail on handling of missing data and sensitivity analyses. Add reference to ICH E9 guidelines."

4 DISCUSSION

4.1 Key Insights

Full automation delivers dramatic time savings while maintaining rigorous standards. Execution automation reduces resource requirements by 50%+, human expertise shifts to strategic oversight. Contrary to initial concerns, this transformation has led to more rigorous quality control, not less, through comprehensive validation frameworks and automated consistency checking.

4.2 Strategic Benefits

- Rapid prototyping of submissions
- Consistent documentation
- Resource optimization
- Complete audit trails
- Novel insight generation

5 CONCLUSION

Viable, wise, valuable? Yes, yes, yes: viability through successful automation of a complete HTA, wisdom through enhanced quality controls and audit trails, and value through dramatic efficiency gains: 5-minute first drafts and 50+% faster projects. For organizations ready to transform their evidence synthesis, the technology is ready to deliver measurable advantages in speed, consistency, and quality today.

ACKNOWLEDGEMENTS: We thank Sophie Wintrich for her valuable input on patient perspectives to this work, and Kerry Hu for her assistance with literature searches.

6 REFERENCES

[1]. https://www.eunethta.eu/ptja08/