Synergistic Polypharmacy Optimization Using Multi-Agent AI and Structured Decision Support

Adnan KULENOVIC [a, 1] and Azra LAGUMDZIJA-KULENOVIC [a]
[a] Absolute Information Age, Inc., Toronto, Canada
ORCID IDs: https://orcid.org/0000-0001-7869-3693, https://orcid.org/0000-0003-1304-9201

Abstract. Polypharmacy management requires both clinical reasoning across multiple conditions and systematic evaluation of adverse drug reactions and interactions among treatment components. This study investigates a synergistic approach that integrates a multi-agent system (MAS) of specialist large language model (LLM) agents with PM-TOM (Personalized Medicine–Treatment Optimization Method), a structured polypharmacy analysis and optimization system. The study hypothesis is that combining agentic clinical reasoning, structured analysis and optimization, and human-guided refinement can identify safer candidate polypharmacy regimens while compensating for key limitations of each approach. This study was based on a published deprescribing case of a 79-year-old man with multiple conditions and medication-related problems. We compared the published discharge treatment (V1) with an AI agent-generated replacement treatment (V2), a PM-TOM-optimized treatment selected from AI-generated drug alternatives (V3), and an AI-assisted clinically refined treatment (V4). PM-TOM evaluated patient-specific drug–adverse drug reactions (ADR), drug–drug interactions (DDIs), drug–condition interactions (DCIs), and Beers Criteria risks. Compared with V1, V4 reduced ADR associations from 38 to 27, critical DDIs from 11 to 4, DCIs from 10 to 7, and combined Beers risk indicators from 9 to 2. The study shows how agentic clinical reasoning, deterministic analysis and optimization, and human judgment can complement each other to improve the safety of complex polypharmacy treatments.

Keywords. Polypharmacy, Clinical Decision Support, Multi-Agent Systems, Large Language Models, PM-TOM, Adverse Drug Reactions, Beers Criteria.

1. Introduction

Polypharmacy is increasingly common in older adults with multimorbidity and is associated with adverse drug reactions (ADRs), drug interactions, potentially inappropriate medications (PIMs), hospitalization, and mortality. A systematic review and meta-analysis of 24 cohort studies involving 2,967,952 adults aged 65 years or older reported a substantial association between polypharmacy and mortality: use of five or more drugs was associated with a 28% higher mortality risk. In comparison, excessive polypharmacy of ten or more drugs was associated with a 44% higher risk [1].

Clinical decision-support systems can identify many medication-related risks, but may produce many alerts without actionable treatment alternatives. More comprehensive

[1] Corresponding Author, Adnan Kulenovic, Absolute Information Age, Inc. Toronto, Canada
E-mail: adnan.kulenovic@abs-info-age.com.

computational approaches are therefore being investigated. A recently validated Artificial Pharmacology Intelligence system, for example, achieved 93.5% agreement with an expert judging committee in identifying medication-related problems in 49 retrospective patient records, illustrating the potential of computational medication review [2]. A 2026 scoping review of 81 studies of AI and rule-based decision support for medication optimization in older adults found consistent improvements in PIM identification, while evidence for patient-level outcomes remained limited [3].

Large language models (LLMs) offer a different capability. Their broad knowledge and natural-language reasoning can integrate clinical context, generate alternatives, explain concerns, and support dialogue. Rao et al. demonstrated ChatGPT's potential for polypharmacy management using standardized deprescribing vignettes, but also reported variability in recommendations by clinical context and across repeated evaluations [4]. LLMs consequently provide flexibility and contextual reasoning but do not inherently perform deterministic combinatorial optimization across treatment alternatives.

Our previous work introduced PM-TOM (Personalized Medicine–Treatment Optimization Method) for patient-specific analysis and optimization of polypharmacy treatments by minimizing cumulative ADRs, drug–drug interactions, and drug–condition interactions [5]. A subsequent case study integrated PM-TOM with a single ChatGPT agent to reduce STOPP and Beers Criteria risks in PIM treatments [6].

The present study extends this approach by using specialist LLM agents as a multi-agent system (MAS). The research hypothesis is that combining agentic clinical reasoning, structured analysis and optimization, and human-guided refinement can find safer polypharmacy regimens while compensating for limitations of each approach.

2. Method and Results

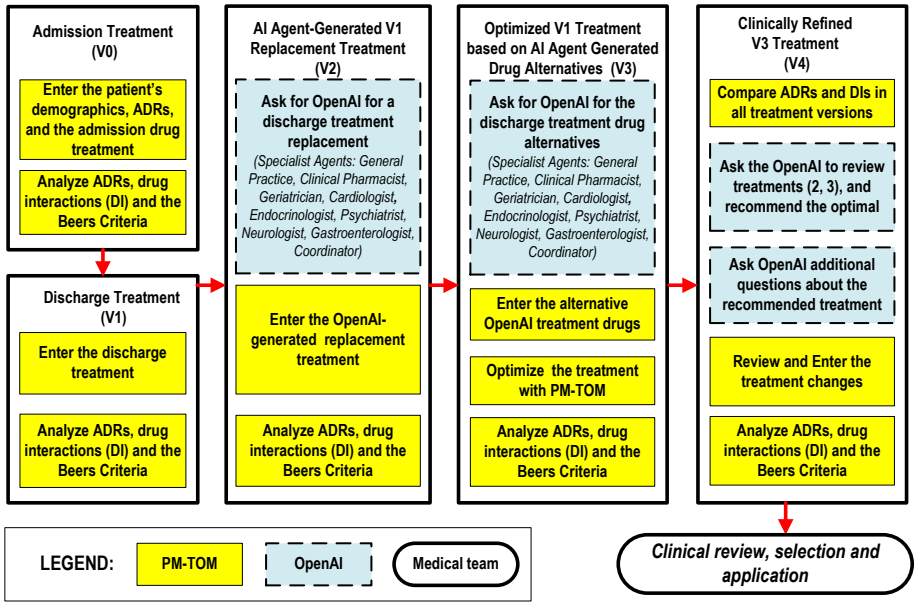
The study was based on the published case of Farrell, Shamji and Dalton involving a 79-year-old man referred to a geriatric hospital for cognition, pain, falls, impaired mobility, mood, dizziness and insomnia [7]. The clinical team performed a comprehensive medication review and produced a modified discharge treatment. This published discharge regimen was designated V1 and served as the reference treatment for the present study. The admission treatment (V0) was retained to document the original case. However, it was not used as the primary comparator because this study did not evaluate the clinical team's deprescribing decisions.

Fig. 1 shows the experimental process. Patient data, V1 treatment, ADRs/symptoms, and other relevant information were supplied to specialist LLM agents representing general practice, clinical pharmacy, geriatrics, cardiology, endocrinology, psychiatry, neurology and gastroenterology. Each was instructed to assess the case within its specialty, consider geriatric prescribing requirements, and identify clinically appropriate recommendations. A coordinator consolidated the specialist reports and preserved issues requiring clinician judgment rather than resolving clinical disagreements independently.

The process produced three V1 modifications. First, the specialist agents proposed a replacement treatment, producing V2. Second, agents generated several clinically appropriate drug alternatives for each treatment part. PM-TOM evaluated combinations of these alternatives and selected V3 by minimizing its Cumulative Adverse Drug Interaction Indicator (CADI). The CADI calculation used the following weights: patient-specific ADR=10, critical DDI=5, other DDI=2, DCI=5, avoid-drug Beers risk=20,

caution-drug Beers risk=10, Beers DDI=20 and Beers DCI=20. PM-TOM uses heuristic search to efficiently identify improved combinations [5, 6].

Fig. 1. MAS-assisted and PM-TOM polypharmacy analysis and optimization.



Finally, V3 underwent an additional AI-assisted clinical review. This step examined alternatives the specialist agents might have missed, including less frequently used but clinically appropriate drugs. PM-TOM then analyzed the resulting candidate alternatives, producing the clinically refined V4. Table 1 summarizes the PM-TOM analysis. Measures included associations between treatment drugs and the patient's ADRs/symptoms; critical and other DDIs; DCIs; and four categories of Beers Criteria risks: drugs to avoid (cADE), drugs to use with caution (oADE), DDIs to avoid and drug–condition interactions to avoid. The accompanying study dataset [8] provides detailed treatment compositions, specialist reviews, analytical reports, and comparison results.

Table 1. PM-TOM comparison of the published and generated treatment versions.

	ADRs and DIs				*The Beers Criteria*			
Version	**ADR**	**cDDI**	**oDDI**	**DCI**	**cADE**	**oADE**	**DDI**	**DCI**
V0 Admission	32	11	42	7	4	1	0	3
V1 Discharge	38	11	40	10	3	2	0	4
V2 MAS replacement	26	7	59	10	1	3	0	3
V3 PMTOM optimized	27	7	42	10	1	1	0	2
V4 Clinically refined	27	4	44	7	1	0	0	1

The published V1 treatment contained 38 patient-specific drug–ADR associations, 11 critical DDIs, 40 other DDIs, and 10 DCIs. Its four Beers categories contained 3, 2, 0 and 4 risks, respectively, for a combined count of 9.

The results show different strengths and limitations at successive stages. V2, generated directly by the MAS, reduced ADR associations from 38 to 26 and critical DDIs from 11 to 7, but other DDIs increased from 40 to 59. Thus, an apparently improved AI-generated regimen did not uniformly reduce interaction risks.

Compared with V1, V3 reduced critical DDIs from 11 to 7, but increased other DDIs from 40 to 42. Further AI-assisted review produced V4, with 27 ADR associations, 4 critical DDIs, 44 other DDIs, and 7 DCIs. Relative to V1, this represents reductions of 29% in ADR associations, 64% in critical DDIs, and 30% in DCIs. The Beers Criteria results showed particularly consistent improvement. The Beers risks total declined from 9 in V1 to 7 in V2, 4 in V3 and 2 in V4, a 78% decrease from V1 to V4.

3. Discussion

The results illustrate the complementary capabilities of generative AI and structured polypharmacy optimization. The specialist agents can interpret multimorbidity, generate treatment alternatives and provide contextual clinical reasoning. PM-TOM instead performs deterministic, reproducible analysis of candidate combinations against structured patient-specific medication-safety data. The MAS expands and reasons about the candidate space; PM-TOM quantitatively analyzes and searches that space; additional AI review can identify overlooked alternatives; and the clinician remains responsible for interpreting the resulting options.

The increase in several warning counts from V0 to the published V1 treatment should not be interpreted as evidence against the original deprescribing team's decisions. The clinical team addressed the patient's broader clinical needs, including introducing two drugs for pain treatment and an additional drug for depression. Adding therapeutically necessary medications can increase the number of potential ADRs and drug interactions, even when clinical treatment improves. Accordingly, the present study evaluates V2–V4 against V1 and does not attempt to retrospectively judge the V0→V1 clinical decisions.

The V2 results illustrate a limitation of unconstrained generative treatment replacement. Although V2 substantially reduced ADR associations, it increased other DDIs. Conversely, structured optimization is constrained by the quality and completeness of its candidate set. During further review of V3, pitavastatin illustrated a potentially useful, less frequently proposed alternative that the specialist agents had overlooked despite instructions to consider less-common and lower-interaction drugs.

A further experiment demonstrated prompt sensitivity in LLM treatment comparison. When OpenAI was initially asked to compare V3 and V4, it selected V3 and relied heavily on established drug use, such as sertraline. When explicitly instructed to compare the treatments first using the supplied patient-specific ADR, DDI, DCI, and Beers findings, the same model selected V4. This reversal does not establish that V4 is clinically superior. Rather, it illustrates that LLM recommendations may depend strongly on how evidence is prioritized in the prompt and suggests that structured PM-TOM results can serve as grounding evidence for subsequent LLM reasoning. This observation is consistent with previous reports of variability in LLM deprescribing decisions [4].

The limitations of this study are as follows. This single-case demonstration does not establish clinical efficacy or patient outcomes; the reported ADR and interaction counts represent potential risks rather than observed clinical outcomes. CADI weights reflect the prioritization used in this experiment and require further clinical validation. LLM

outputs are model- and prompt-dependent, while structured optimization depends on the completeness and accuracy of its medication data and candidate alternatives.

4. Conclusions

This case study demonstrates the potential of combining multi-agent LLM reasoning with structured polypharmacy analysis and optimization. Compared with the published discharge treatment, successive refinements reduced patient-specific ADR and interaction indicators, with substantial reductions in critical DDIs and Beers Criteria risks. At the same time, intermediate results exposed limitations of both approaches: LLM agents can overlook useful alternatives or prioritize familiar therapies, whereas structured optimization cannot identify options outside its supplied candidate set or independently determine their broader clinical appropriateness.

The combined method has practical value beyond retrospective deprescribing. Applied in primary care, it could help identify safer alternatives before prescribing unsafe regimens and thus reduce the subsequent need for complex deprescribing interventions. The method could also help pharmacists and other clinicians monitor and adjust polypharmacy treatments. Prioritized recommendations and treatment alternatives may reduce non-actionable alerts and alert fatigue, simplifying the search for safer multidrug regimens, and allowing healthcare professionals to focus on clinical judgment. The intended role is therefore not autonomous prescribing but collaborative decision support in which structured computation, agentic AI and healthcare professionals contribute complementary strengths.

5. References

[1] Li Y, et al. Association between polypharmacy and mortality in older adults: a systematic review and meta-analysis. Arch Gerontol Geriatr. 2022;100:104630. doi: 10.1016/j.archger.2022.104630.

[2] Dil-Nahlieli D, Ben-Yehuda A, Souroujon D, Hyam E, Shafran-Tikvah S. Validation of a novel Artificial Pharmacology Intelligence (API) system for the management of patients with polypharmacy. Res Social Adm Pharm. 2024;20(7):633-639. doi: 10.1016/j.sapharm.2024.04.003.

[3] Eskandar K. Artificial intelligence and rule-based decision support for medication optimization in older adults: a systematic scoping review of polypharmacy management tools and ethical challenges. Geriatric Pharmacology. 2026;3(1):2596797. doi: 10.1080/2994399X.2025.2596797.

[4] Rao A, Kim J, Lie W, et al. Proactive polypharmacy management using large language models: opportunities to enhance geriatric care. J Med Syst. 2024;48(1):41. doi: 10.1007/s10916-024-02058-y.

[5] Kulenovic A, Lagumdzija-Kulenovic A. PM-TOM: a method for finding personalized polypharmacy therapies with minimal adverse drug-drug, drug-gene and drug-condition interactions. Stud Health Technol Inform. 2020;270:648-652. doi: 10.3233/SHTI200240.

[6] Kulenovic A, Lagumdzija-Kulenovic A. Minimizing STOPP and Beers criteria risks in PIM treatments using PM-TOM and ChatGPT: a case study. Stud Health Technol Inform. 2025;323:149-153. doi: 10.3233/SHTI250067.

[7] Farrell B, Shamji S, Dalton D. Managing chronic diseases in the frail elderly: more than just adhering to clinical guidelines. Can Pharm J (Ott). 2014;147(2):89-96. doi: 10.1177/1715163513514021.

[8] Kulenovic A, Lagumdzija-Kulenovic A. Synergistic agentic AI and data-driven optimization for polypharmacy management. Mendeley Data; 2026. doi: 10.17632/mrn5xhvyhb.2.