A Hybrid Methodology to Bridge Expert Insight and AI Reasoning

H. Izadbakhsh[1], M. Zarinbal[2], S.Z Kalantari-Banadaki[3,4], N. Gh.Tehrani[5]

[1] Department of Industrial Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran.
Email: hizadbakhsh@khu.ac.ir,
https://orcid.org/0000-0001-9068-6883

[2] Department of Industrial Engineering and Management Systems, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran
Email: mzarinbal@aut.ac.ir
https://orcid.org/0000-0001-9284-7329

[3] University of Hormozgan, Bandarabbas, Iran.
[4] Statistical Methods and Population Modeling Group, National Institute for Population Research, Tehran, Iran.
Email: kalantarizahra1987@gmail.com
https://orcid.org/0000-0002-3099-4638

[5] Department of Industrial Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran.
Email: nasim.tehrani@gmail.com
https://orcid.org/0000-0002-9662-8235

Abstract

Major public policy challenges such as climate change, population, welfare, and resource scarcity are inherently complex, shaped by nonlinear feedbacks. This paper presents a new hybrid methodology to construct a Causal Loop Diagram (CLD) in a six-step process. In this methodology, Group Model Building (GMB), Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG) have been integrated: (1) formulating a dynamic hypothesis; (2) developing an expert-driven CLD; (3) generating an AI-assisted CLD via few-shot prompting; (4) merging human- and AI-derived models; (5) using RAG to anchor causal links in peer-reviewed evidence; and (6) validating the integrated model through expert review. This methodology was applied on population dynamics and demonstrated that how combining expert insight and AI reasoning can enhance the transparency, scope, and policy relevance of systems modeling in data-limited and fast-changing contexts.

Keywords: Hybrid Methodology; Systems Thinking; Group Model Building; Large Language Models; Retrieval-Augmented Generation.

1 Introduction

A systems perspective is essential to capture the dynamic interactions for public policy challenges. Despite years of research on the drivers and consequences of these challenges, public policies have often focused on isolated variables. This narrow focus overlooks the interdependencies, feedback loops, and nonlinearities that define in dynamic systems (Kunte & Damani, 2015).

Thus, in developing CLD for such challenges the main question could be defined as: *Which reinforcing and balancing feedbacks best explain the problem, and which policy packages could be more effective?*

To answer this question, an integrated systems-thinking workflow is employed that combines Group Model Building (GMB) with large language models (LLMs) and Retrieval-Augmented Generation (RAG). GMB anchors the model in expert knowledge and structured facilitation, while LLMs accelerate hypothesis generation and expand coverage of plausible feedbacks. RAG then grounds candidate causal links in published literature to reduce hallucination and improve traceability. The intent is not to replace expert judgment with automation, but to create a transparent, reproducible human–AI collaboration that yields a clearer, evidence-informed causal narrative.

2 Methods

The framework proceeds from Formulating the Initial Dynamic Hypothesis (Step 1), then moving on to Developing an Expert-Driven CLD (Step 2). Through this, by formulating the human-derived CLD, it ensures that the model is grounded within a particular socio-political environment. Steps 3 and 4 consists of AI-Driven CLD Construction (Step 3) and Merging the Expert- and AI-Derived Models (Step 4). These two steps utilize LLMs potential to identify feedback structures that experts might overlook due to their domain-specific focus or cognitive biases. The final steps focus on CLD Refinement using RAG (Step 5) and Expert Validation (Step 6). To mitigate the risk of AI hallucinations and ensure theoretical alignment, Step 5 employs RAG to anchor causal links in peer-reviewed literature. Finally, returning the refined model to a validation panel (Step 6) ensures that the synthesized human-AI insights remain conceptually clear and policy-relevant.

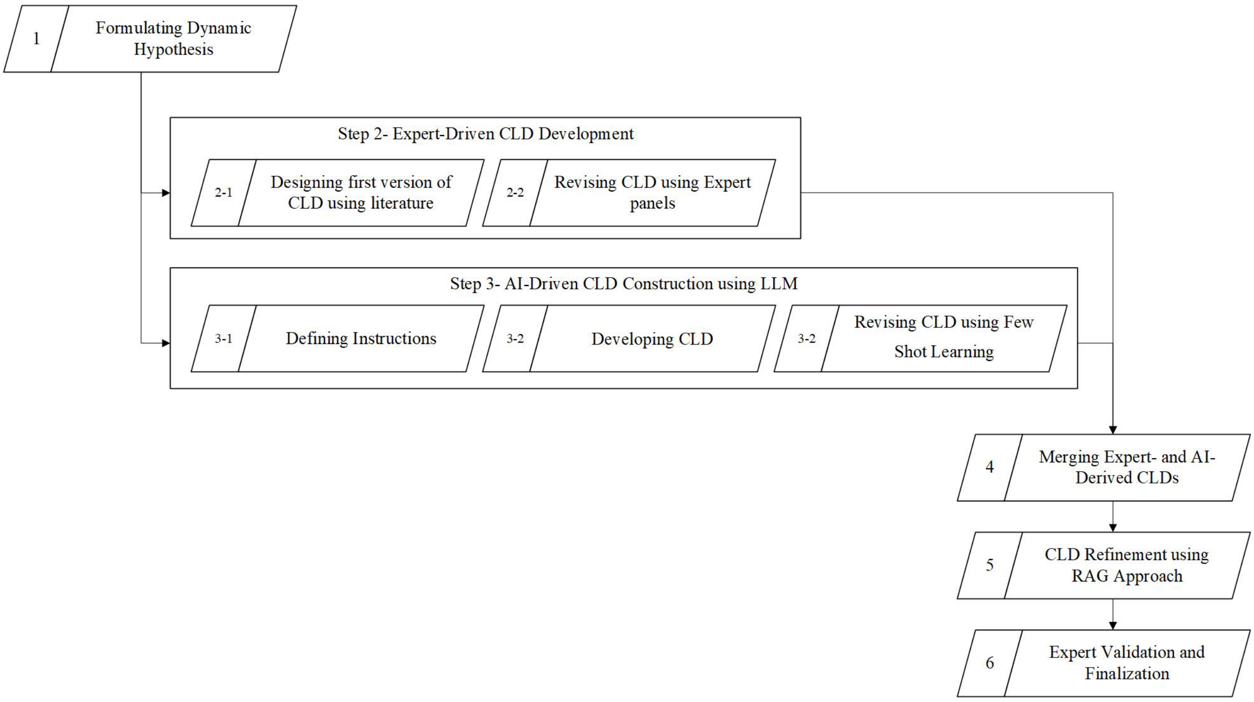

Figure 1- Stepwise Methodology

3 Results

To verify the proposed methodology, population dynamics is selected. Population is the core of major policy challenges in health, welfare, housing, labor markets, and environmental sustainability. Despite years of research on the drivers and consequences of demographic change, population policies have often focused on isolated variables such as fertility, mortality, or migration. Understanding how reinforcing and balancing feedback mechanisms interact in this context is critical for developing effective, cross-sector policy responses.

In the following the development and refinement of the CLD through the six-step framework outlined in the Methods section is presented.

- Step 1- Formulating the Initial Dynamic Hypothesis

The following initial dynamic hypothesis represents a narrative synthesis of the observed demographic shifts. This statement provides the qualitative shape and boundary for the modeling process, which is then expanded and validated in the later steps:

Initial Dynamic Hypothesis
Demographic transformations have been driven by a complex interplay of governance policies, economic development, and socio-cultural change. Earlier fertility declines associated with improved healthcare, urbanization, female education, and effective family-planning programs. These trends progressively shifted the population structure toward aging.

Rising life expectancy, delayed marriage, declining fertility preferences, and the weakening of intergenerational support systems have created reinforcing feedback loops that accelerate population aging. Existing pro-natalist policies have only partially offset these trends. Without structural transformations in economic, cultural, and labor market domains, the countries risk entering a low-fertility and population-aging trap.

- Step 2: Expert-Driven CLD Development

An initial CLD using three focused GMB sessions and one workshop was developed to capture key feedback structures influencing fertility dynamics. This collaborative process ensured multidisciplinary understanding and helped uncover overlooked or contested causal pathways relevant to fertility dynamics.

However, this process also faces several challenges, including cognitive biases (Doyle et al., 2009; Vennix, 1999), difficulty reaching consensus (Király & Miskolczi, 2019), risks of oversimplification (Rouwette & Smeets, 2016), and heavy dependence on facilitator skill. The manual construction of causal loop diagrams (CLDs) is also time-consuming and may omit key variables or relationships (Liu & Keith, 2025).

To address these limitations, recent studies have been conducted to offer practical improvements (Cruden et al., 2024; Deutsch et al., 2024; Keenan et al., 2025; Zarinbal et al., 2025). In parallel, artificial intelligence (AI), especially LLMs, has shown promise in extracting causal relationships from text, generating CLDs, and expanding the participatory modeling space (Hosseinichimeh et al., 2024; Liu & Keith, 2025; Long et al., 2024). Thus, in the next step, OpenAI's ChatGPT-4o is used in to generate an alternative CLD.

- Step 3: AI-Driven CLD Construction using LLM

To have an alternative insight, a state-of-the-art LLM, OpenAI's ChatGPT-4o version May 2024, was used by modelling team to generate an alternative CLD. Given the complexity and domain specificity of the task, few-shot prompting was used to guide the ChatGPT (OpenAI et al., 2024). The model received the following instructions and prompt:

ChatGPT Instructions:

You are a professional scientist and your research focuses on population dynamics. You analyze it through the lens of systems thinking and system dynamics methodologies.

Objective:

Develop a Causal Loop Diagram (CLD) that captures the key drivers of population dynamics, both historically and as a tool for future policy-making. The CLD should incorporate both Reinforcing (positive) and Balancing (negative) feedback loops.

Key Requirements:

- Historical Explanation: The CLD should elucidate past population trends (e.g., demographic transitions, industrialization effects).
- Policy Relevance: The model should serve as a decision-support tool for future population management (e.g., aging societies, sustainability challenges).
- Dynamic Complexity: Capture nonlinear interactions, delays, and emergent behaviors in the system.

Prompt:
Considering the Project’s Instructions and the Initial Dynamic Hypothesis, develop a CLD that captures the key dynamics, both historically and as a tool for future policy-making.

The AI-generated CLD was reviewed by the modelling team. In the generated CLD there were variables and directions which were ambiguous, and some loops were not closed form or an actual loop. Therefore, ChatGPT was asked to modify or explain the rationale.

- Step 4: Merging Expert- and AI-Derived CLDs

The comparison of the Expert and AI Derived CLDs illustrates two fundamentally different ways of conceptualizing population dynamics. The Expert Derived CLD models population as a complex system influenced by experts’ knowledge. Conversely, the AI Derived CLD recognizes that patterns of fertility are relatively consistent across different cultures and societies and will be driven by globalization and modernization, rather than being the direct result of government decisions. When viewed together, these two perspectives don't compete, but rather complement one another. The AI Derived Model assists in identifying the significant forces, while the Expert Derived Model illustrates how those forces interact within the context through social and policy environments. Thus, to benefit from these perspectives, in step 4, expert and AI-generated models were merged, resulting in a more comprehensive system representation.

- Step 5: CLD Refinement using RAG

To strengthen the empirical foundation of the integrated CLD, a Retrieval-Augmented Generation (RAG) framework was implemented. Typically, standard LLMs create responses based only on their internal training datasets, while RAG allows the use of external source documents to help the model create its response. This ensures that the model’s modifications to the CLD are grounded in verifiable empirical evidence rather than generic generative patterns. That is, applying RAG, represents a methodological shift from generative AI to verifiable AI. This process reduces the incidence of hallucination, or incorrect information, and provides theoretical basis (Lewis et al., 2020).

The RAG corpus comprised of 28 peer-reviewed journal articles, 6 statistical or institutional reports, 3 book chapters, and academic theses and articles, with deliberate emphasis on demographic, socioeconomic, and policy context.

When prompted, ChatGPT performed a semantic search across the corpus to identify relevant passages. It was instructed to read the attached sources and, if necessary, make some recommendations (edit, add or delete the variables or loops). The recommendations should include the rationale and the relevant paragraph from sources. The LLM then refined the existing CLD structure based *only* on the retrieved evidence, providing the rationale for every modification.

In this process, demographic theories (Second Demographic Transition (SDT), proximate determinants of fertility, and low-fertility trap hypotheses was systematically map into endogenous feedback structures. This ensured that the model was not only comprehensive but also grounded in verified sources, reducing hallucination risks commonly associated with generative AI models.

- Step 6: Expert Validation and Finalization

The finalized CLD was reviewed by a validation panel comprising specialists in system dynamics, demography, and public policy. The panel assessed the conceptual clarity, empirical plausibility, and policy relevance of each loop and subsystem. Their feedback led to informed refinements in loop articulation, variable definitions, and subsystem boundaries. The final version reflects a high degree of conceptual clarity, and empirical plausibility.

Figure 2 presents the finalized CLD.

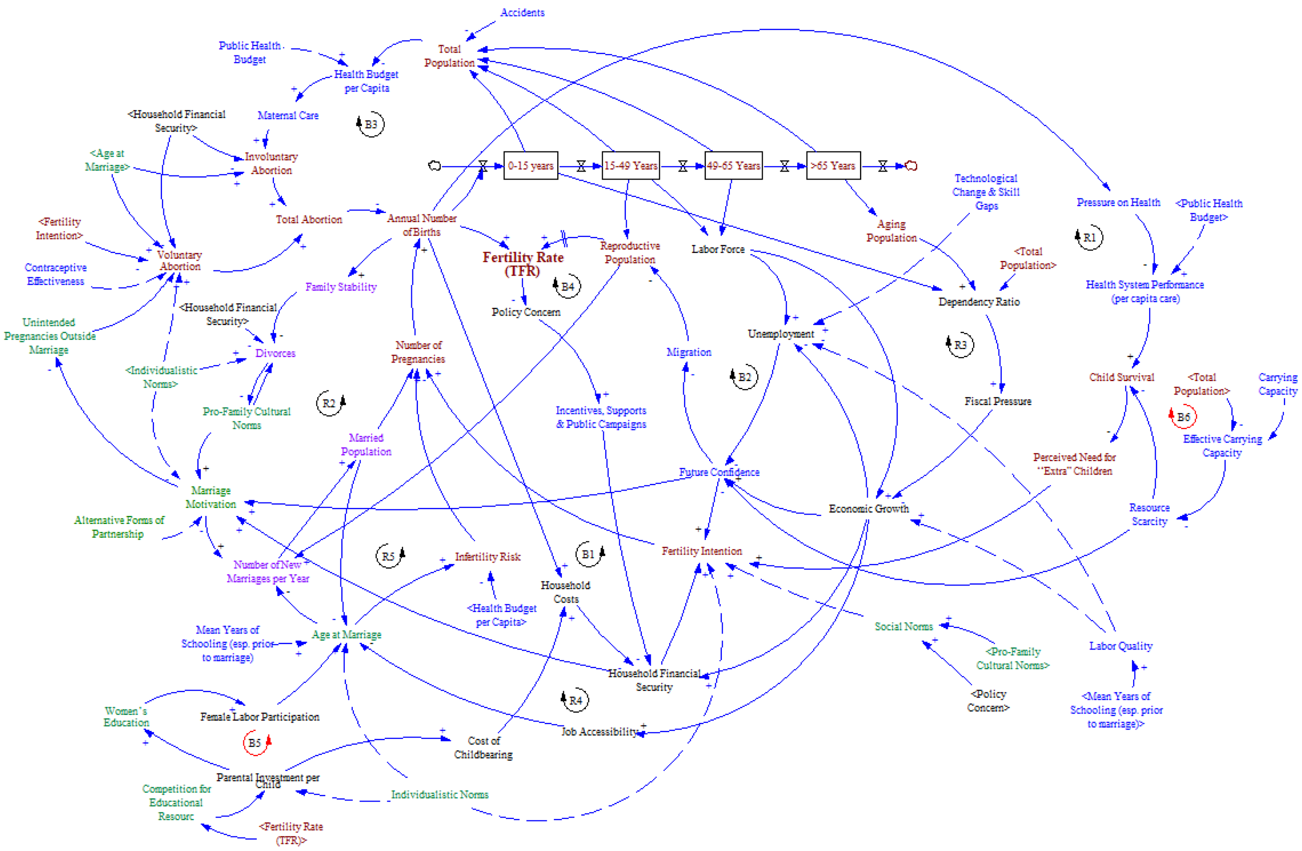

Figure 2- Final CLD of Population Dynamics

System-dynamics experts recommend that an aging-chain structure should be added to reflect demographic momentum. So, the final CLD incorporates an aging-chain structure with four cohorts: 0–15 years, 15–49 years (reproductive population), 49–65 years, and 65+ years. Population flows occur through births (Annual Number of Births) and aging transitions from one cohort to the next. By including these cohorts, the model explicitly accounts for the "time-to-age" delay. This ensures that the model reflects how today's birth rates dictate the size of the reproductive population two decades into the future, critical feedback that a simple variable-based CLD might oversimplify.

The stronger presence of balancing loops reflects broader trends in demographic literature, where scholars tend to emphasize stabilizing constraints, such as economic pressure, delayed marriage, and health limitations, over amplifying dynamics.

4 Discussion

This study produced two primary outcomes: (1) methodological lessons from applying LLMs in a hybrid participatory modeling process, and (2) policy insights derived from the integrated model of population dynamics.

- **Potentials and Limitations of Applying AI**

Combining expert knowledge with LLM enriched the causal structure of the final CLD. The expert-derived model offered deep contextual insight into fertility, family, and economic dynamics, while the AI-generated model contributed complementary feedback structures related to education, urbanization, and modernization. This integration allowed the model to bridge local insight and global demographic theory and enhanced the depth and relevance of the final model.

However, this experience also exposed the limitations of LLMs in complex system modeling. Empirical studies show that LLMs often overlook key causal links and yield outputs sensitive to prompt phrasing. For example, the "System Dynamics Bot" by (Hosseinichimeh et al., 2024) identified only about 60% of causal links when compared with expert-developed models. (Long et al., 2024) also showed that LLMs' suggestions change depending on query phrasing, reducing reliability without expert oversight. Moreover, hallucination is another risk; LLMs may suggest causal relationships that do not exist, especially in specialized domains.

To mitigate these weaknesses, this study implemented two safeguards:

- Applied RAG to anchor LLM outputs in peer-reviewed research: What RAG can provide in formulating a qualitative model would be its capacity to trace the mapping of demographic theories, into endogenous feedback structures based on the provided corpus. In linking AI results to the peer-reviewed literature, LLMs could be transformed from a creative brainstorming tool to a precise tool. This would rectify the hallucination problem through tracing each link in the CLD to a textual citation in the literature.

  In addition, RAG enables a more comprehensive assessment of balancing loops that are typically overlooked or only marginally addressed in GMB.
- Few-shot prompting to ensure domain relevance and to correct causal polarities: By providing feedbacks, the modeling team anchored GPT's focus on demographic theory rather than general trends, and thus showed the model exactly what was expected. Moreover, complex tasks like modeling population dynamics require Chain-of-Thought (CoT) reasoning. By pairing few-shot examples with instructions to "show the rationale," we forced GPT to break down its reasoning into intermediate steps before determining the final loop structure.

In short, AI helped accelerate model development and broaden the perspective, but expert judgment was essential to ensure relevance and validity.

- **Policy Implications**

The integrated CLD reveals that population dynamics are systemic, multidimensional, and policy-sensitive. Fertility decline, in particular, is influenced not only by demographic trends but also by labor market structures, education, healthcare capacity, family norms, and public policy. A systems thinking perspective shows that population policies must address these domains in an integrated way to avoid fragmented and counterproductive results. These policies must be adaptive and cross-sectoral. Narrow, fertility-focused approaches often fail when disconnected from economic, health, housing, or education systems (Joulaei et al., 2025).

5 Conclusion

This paper presents a comprehensive, innovative methodology to model major public policy challenges. This methodology uses AI not to replace the experts, but as a reasoning tool. Few-shot prompting was used to give LLM specific instructions to help anchor its focus and ensure domain relevance and logical consistency. Furthermore, using RAG allowed the model to anchor its outputs in a corpus consist of global and regional demographic literature. This process transformed the AI from a generative tool into a verifiable tool and uncovered critical balancing loops that were absent in initial CLDs. Together, the human-machine integration produced a more coherent, empirically grounded, and policy-relevant CLD than either alone could achieve.

The proposed modeling methodology was applied to population dynamics but it offers options for other complex domains where progress rests on the interactions among constituent subsystems and feedback. It is especially effective in contexts with limited or rapidly changing data where it may be important for diverse perspectives to be captured in timely manner.

Funding

This work was supported by the Iran National Science Foundation (4031584).

References

Cruden, G., Powell, B. J., Frerichs, L., Lanier, P., Brown, C. H., Saldana, L., & Lich, K. H. (2024). Leveraging group model building to operationalize implementation strategies across implementation phases: an exemplar related to child maltreatment intervention selection. *Implementation Science Communications*, *5*(1), 134. https://doi.org/10.1186/s43058-024-00660-2

Deutsch, A. R., Frerichs, L., Perry, M., & Jalali, M. S. (2024). Participatory modeling for high complexity, multi-system issues: challenges and recommendations for balancing qualitative understanding and quantitative questions. *System Dynamics Review*, *40*(4). https://doi.org/10.1002/sdr.1765

Doyle, J. K., Ford, D. N., Radzicki, M. J., & Trees, W. S. (2009). Mental Models of Dynamic Systems. In Y. Barlas (Ed.), *Systems Dynamic* (Vol. 2, 11). EOLSS Publications. https://www.eolss.net/ebooklib/bookinfo/system-dynamics.aspx

Hosseinichimeh, N., Majumdar, A., Williams, R., & Ghaffarzadegan, N. (2024). From text to map: a system dynamics bot for constructing causal loop diagrams. *System Dynamics Review*, *40*(3), e1782. https://doi.org/https://doi.org/10.1002/sdr.1782

Joulaei, H., Morshed-Behbahani, B., Ghadimi, P., Nadimi Parashkouhi, S., & Mansoori, Y. (2025). Contentious population policy-making and its consequences: a health policy analysis. *International Journal for Equity in Health*, *24*(1), 96. https://doi.org/10.1186/s12939-025-02469-4

Keenan, M., Freeman, L., Santana de Lima, E., Potter, K., Hobbs, T., Ballard, E., & Fonagy, P. (2025). A systemic approach to identifying sustainable community-based interventions for improving adolescent mental health: a participatory group model building and design protocol. *Health Research Policy and Systems*, *23*(1), 9. https://doi.org/10.1186/s12961-024-01247-y

Király, G., & Miskolczi, P. (2019). Dynamics of participation: System dynamics and participation—An empirical review. *Systems Research and Behavioral Science*, *36*(2), 199–210. https://doi.org/10.1002/sres.2580

Kunte, S., & Damani, O. P. (2015). Population Projection for India–A System Dynamics Approach. *33rd International Conference of the System Dynamics Society*, 19–23.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *34th International Conference on Neural Information Processing Systems (NIPS'20)*, 9459–9474.

Liu, N.-Y. G., & Keith, D. R. (2025). Leveraging Large Language Models for Automated Causal Loop Diagram Generation: Enhancing System Dynamics Modeling through Curated Prompting Techniques. *ArXiv*.

Long, S., Schuster, T., & Piché, A. (2024). Can large language models build causal graphs? *ArXiv*.

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2024). GPT-4 Technical Report. *ArXiv*.

Rouwette, E., & Smeets, S. (2016). Conflict, consensus and the management of a good debate. Exploring the deliberative assumptions of group facilitating techniques. In I. Bleijenbergh, H. Korzilius, & E. Rouwette (Eds.), *Methods, Model Building and Management* (pp. 129–146). Institute for Management Research.

Vennix, J. A. M. (1999). Group model-building: tackling messy problems. *System Dynamics Review*, *15*(4), 379–401. https://doi.org/10.1002/(SICI)1099-1727(199924)15:4<379::AID-SDR179>3.0.CO;2-E

Zarinbal, M., Izadbakhsh, H., & Tehrani, N. G. (2025). People empowerment in the healthcare system: A conceptual model and roadmap. *Public Health*, *243*, 105720. https://doi.org/https://doi.org/10.1016/j.puhe.2025.03.032