Artificial Intelligence versus Researcher: Conceptual Model Development

Elliott Blatt,[1] Lara Sams,[2] Conrad Bessant,[3] Maryam Abdollahyan,[3] Fabrizio Smeraldi,[3] Tara Symonds[2]

[1] Clinical Outcomes Solutions, Tucson, AZ, US; [2] Clinical Outcomes Solutions, Folkestone, Kent, UK; [3] Mebomine Ltd., London, UK

Poster 2077

Background

- Conceptual model (CM) development plays a vital role in understanding complex medical conditions such as Lennox Gastaut Syndrome (LGS).
- A CM can be used to represent patient's specific health experiences and to visualize the concepts that describe those experiences. This can be useful to the FDA and sponsors when determining whether a proposed clinical outcome assessment (COA) measure sufficiently captures a concept of interest and measures what is important to patients.[1]
- Traditionally, CMs are developed through conducting literature reviews, a relatively time-consuming process for researchers.
- With the advent of artificial intelligence (AI), methods have emerged as potential tools for automating CM development, for example, use of chatbots that use pretrained large language models to generate conversations (e.g., ChatGPT), and the use of AI to mine online health forums for data (e.g., Mebomine).

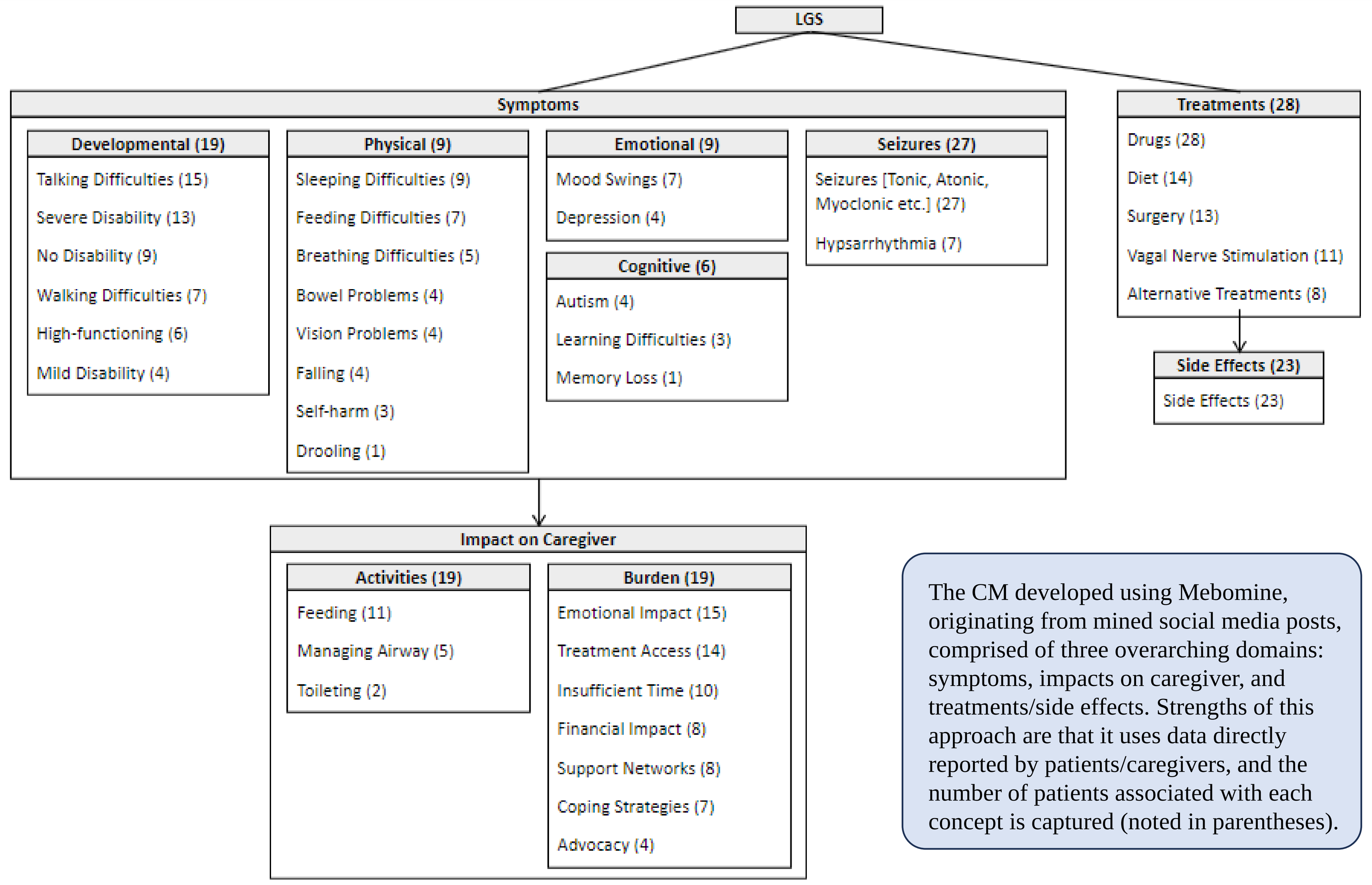
Objectives

- The objective of this study was to compare the process and outcomes of CM development between researchers, ChatGPT v3.5, and Mebomine's v23.1 analysis platform.

Methodology

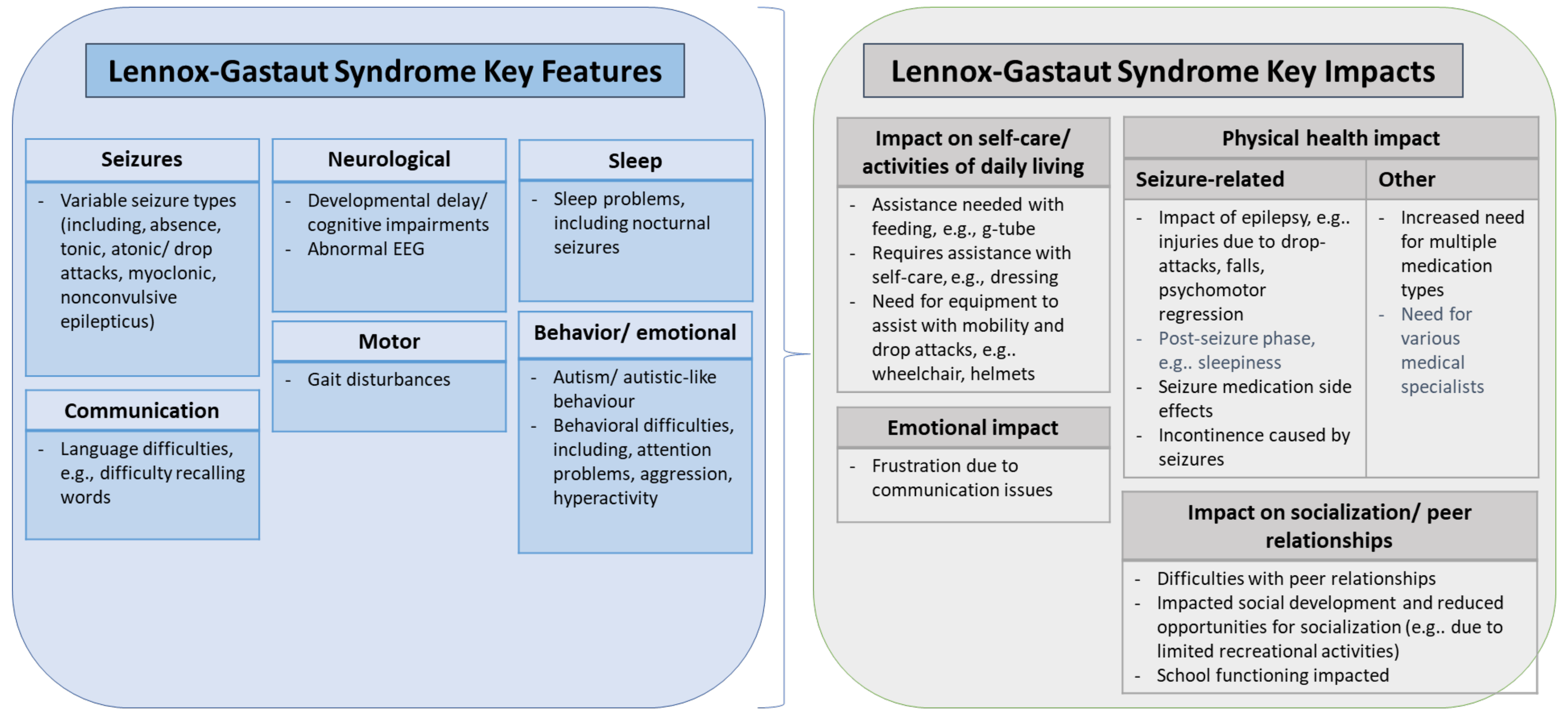
- Several methodologies were utilized to generate a CM to represent the key concepts of LGS:
  - Researchers conducted a literature review utilizing the Ovid database (Embase, MEDLINE, and PsycINFO) to identify articles detailing key symptoms and impacts of LGS. Researchers then reviewed the articles to construct a LGS disease CM.
    - 167 articles were identified. Title, abstract, and full text reviews were then conducted by researchers to exclude articles.
  - ChatGPT was used to generate a CM based on input prompts. ChatGPT does not search the internet for information but instead uses information learned from training data.
  - Mebomine used the condition term to conduct a search of online heath boards (OHBs) using an in-house search engine. The content of these was analyzed using human-guided natural-language processing to identify the concepts being discussed and determine basic patient demographics.
    - Mebomine retrieved 675 English language posts from OHBs. These were authored by 41 caregivers of those with LGS.
- These methodologies and the CM generated by each were compared by researchers to determine the differences between concepts.

Figure 1 Lennox Gastaut Syndrome (LGS) Conceptual Model – Developed by Mebomine



The CM developed using Mebomine, originating from mined social media posts, comprised of three overarching domains: symptoms, impacts on caregiver, and treatments/side effects. Strengths of this approach are that it uses data directly reported by patients/caregivers, and the number of patients associated with each concept is captured (noted in parentheses).

Figure 2 Lennox Gastaut Syndrome (LGS) Conceptual Model – Developed by Researcher

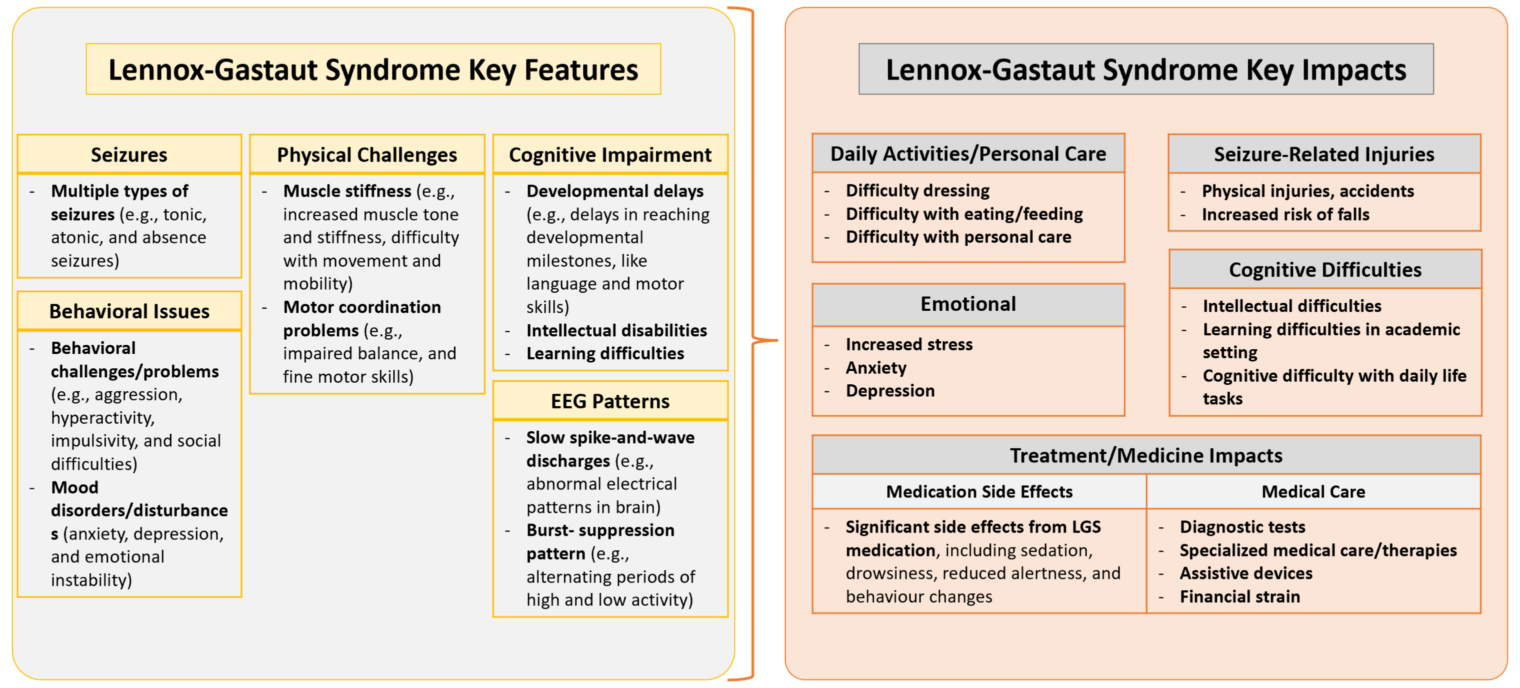


The CM developed by a researcher using published literature is comprehensive, however, limited with respect to any previously published literature. For this CM, information was gained from a small number of articles noting symptom and impacts in LGS. In many cases, patient-centered or -reported literature on a particular condition, especially rare diseases, can be limited.

Results

- Our findings indicated that all 3 methods produced a reasonable summary of the key features and impacts to assist in an understanding of the disease presentation.
- General concepts tended to align across the CMs, with similar high-level domains identified (e.g., seizures, physical functioning/ motor, and behavioral/emotional functioning) and multiple concepts grouped under each domain (see Figures 1, 2 and 3).
- Frequencies of reporting of concepts was provided by both the researcher led CM and the Mebomine CM, but no frequencies were able to be tracked for the CM developed using ChatGPT.
- In addition to this, both the Mebomine and researcher led literature review CM (Figure 1 and 2) allowed the source data to be accessed at a later point or for further details to be retrieved. This is not the case with ChatGPT, where the source data is unclear, and how decisions were made to include or exclude certain concepts are unknown. Currently (September 2023) and as stated within the ChatGPT output, the ChatGPT training data only goes up until September 2021, which means information from after this time would not be included.
- A key finding of the ChatGPT CM (Figure 3) identified was the inconsistent results when inputting the same prompts on different occasions.
  - The way that categories were divided differed, as well as the features listed, and short descriptions provided. Models generated from ChatGPT therefore looked different upon running several times. For example:
    - On one occasion it categorized by 'seizure type', 'cognitive', 'motor', etc., whereas on another occasion it categorized by 'key features', 'symptoms', and 'impacts'.
    - On one occasion it included symptoms such as 'mood disorders', 'refractory epilepsy' and 'status epilepticus', whereas on another occasion it did not include these symptoms within the output.

Figure 3 Lennox Gastaut Syndrome (LGS) Conceptual Model – Developed using ChatGPT

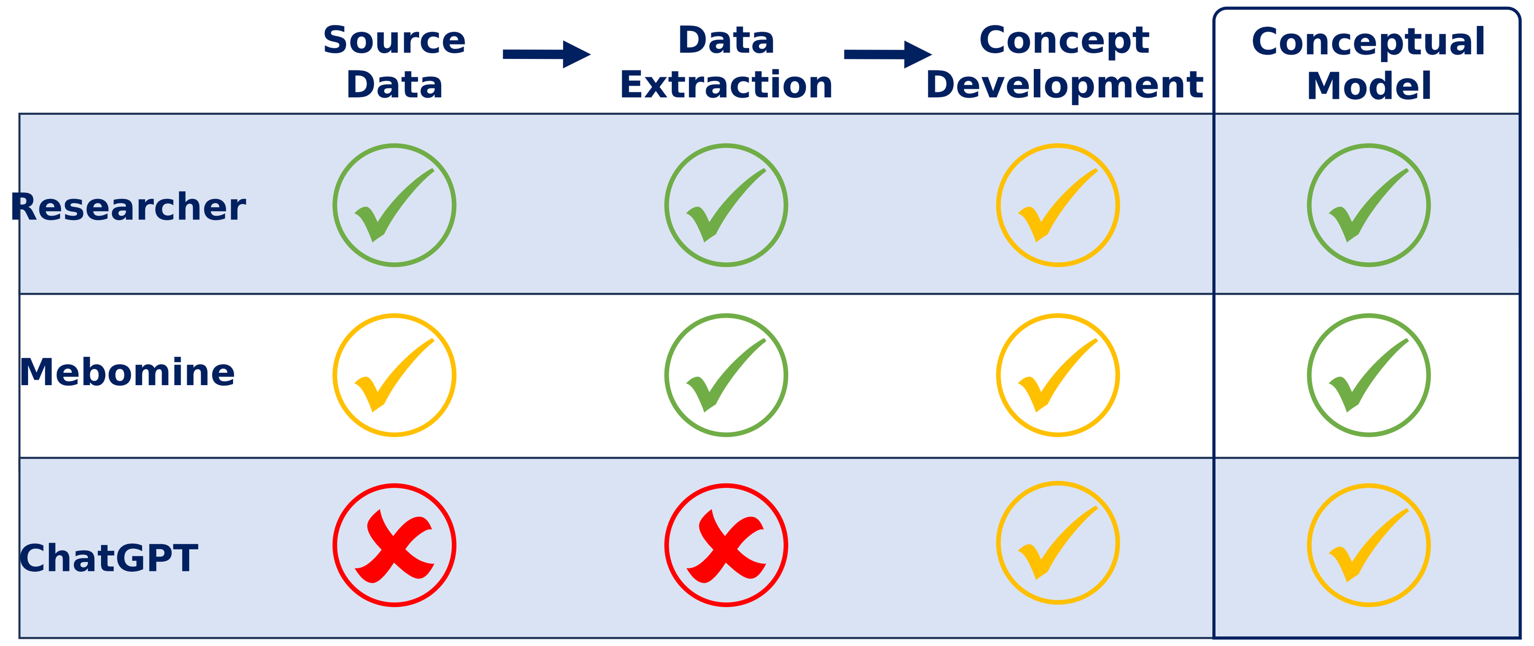


The CM developed by ChatGPT is a near match to the concepts that were derived from literature by researchers. A limitation of this method is that the source data is unknown, therefore, it is likely that ChatGPT is simply predicting what the researcher would want. There is significant potential for error in this source data.

Results (cont.)

- The searching of OHBs (via Mebomine) is restricted to key health areas that the patient or caregiver is focused upon during the time of writing their post. For example, they may focus posts upon depression (mentioned by n=4 caregivers), as this is a key concern for them and the reason for creating a post.
  - For example, of 41 caregivers, 27 mentioned seizures within the OHBs reviewed. LGS is characterized by multiple seizure types including tonic, atonic, atypical absence, and generalized tonic-clonic seizures[2] and thus it is likely that a higher proportion of these patients present with seizures than those who mention this concept within the OHBs.
  - This is something that would need to be considered when working to ensure that the CMs developed capture the wider disease experience.

Figure 4 Evaluation of Confidence in Conceptual Model Development Process



= High confidence
= Medium confidence
= Low confidence

Each methodology was evaluated for quality and reproducibility at each step in the CM development process. Of note, for the concept development stage, all methods were assigned a level of medium confidence attributed to any potential for nuances between the steps between data collection and assembling concepts into the CM.

Conclusions

- Both ChatGPT and Mebomine produced detailed CMs that were similar to the researcher-led CM, and so can offer advantages in terms of automation and efficiency.
- Although Mebomine produced a CM with key concepts identified, further work would be needed by researchers to clearly delineate and contextualize these concepts e.g., for developmental delay it was described across the severity spectrum but would be best listed as 'variable disability'.
- Where possible, literature reviews should obtain data from qualitative studies to ensure the patient voice is promoted throughout the CM development. Importantly, researchers bring unique human insights and contextual understanding but, there is also a subjective element to researcher led literature reviews that should be acknowledged, for example, categories having the potential to be grouped in different ways, or concepts being described differently.
- A collaborative approach that combines the strengths of both human researchers and AI may offer the best path forward, allowing for efficient automation while benefiting from the human expertise needed for CM development.
- The incorporation of data from OHBs can also ensure promotion of the patient voice, particularly in cases where published qualitative articles are limited. However, further literature should be reviewed, when possible, to ensure that the full range of symptoms are captured.
- Being able to retrieve the source data is an important factor in the ability to reproduce the results and ensure that results are accurate and therefore, further research into the integration of AI into medical research is needed before confidence can be instilled in the results of solely AI-generated work.
- Following the initial drafting of a disease CM, work to further capture the patient voice in the CMs is recommended. Ideally this would be through patient or caregiver interviews to ensure that the most important concepts are captured and that these are well-defined within the model.

References

1. U.S. Food & Drug Administration. Patient-Focused Drug Development: Selecting, Developing, or Modifying Fit-For-Purpose Clinical Outcome Assessments. FDA Website. June 2022.
2. Al-Banji MH, Zahr DK, Jan MM. Lennox-Gastaut syndrome. Management update. Neurosciences (Riyadh). 2015;20(3):207-212. doi:10.17712/nsj.2015.3.20140677

Poster presented at the ISOQOL 30th Annual Conference, October 18-21, 2023, Calgary, Canada

Acknowledgments Takeda Pharmaceutical Company Limited, Cambridge, MA provided permission to use a portion of these data for this study. Editorial support was provided by Clinical Outcomes Solutions.