

Development of a Core Outcome Set for Neurological Disorders (COS-Neuro): an AI-Assisted Thematic Framework Analysis.

Shyun Ping Tiong, Xiaoyu Yang, Alvaro Yanez Touzet, Christopher Paul Millward, Carl M Zipser, Lindsay Tetreault, Ali Gharooni, Benjamin Davies

Submitted to: Journal of Medical Internet Research
on: April 23, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 36

0..... 36

..... 36

Preprint
JMIR Publications

Development of a Core Outcome Set for Neurological Disorders (COS-Neuro): an AI-Assisted Thematic Framework Analysis.

Shyun Ping Tiong^{1, 2} MD; Xiaoyu Yang^{3, 4} MD, MSc, PhD; Alvaro Yanez Touzet⁵ BSc; Christopher Paul Millward⁶ MBBS, PhD; Carl M Zipser^{7, 8} MD, PD Dr. med, FEBN, M.A.; Lindsay Tetreault⁹ MD, PhD; Ali Gharooni^{10, 11, 12} MBChB, MSc, MRCS; Benjamin Davies^{10, 13} BSc, MBChB, MPhil, PhD

¹ Queen Alexandra Hospital Portsmouth University Hospitals NHS Trust Portsmouth GB

² Wessex Deanery, Health Education England Portsmouth GB

³ Imperial College London London GB

⁴ Imperial College Healthcare NHS Trust London GB

⁵ University of Manchester Manchester GB

⁶ University of Liverpool Liverpool GB

⁷ Universitätsklinik Balgrist Zurich CH

⁸ University of Zurich Zurich CH

⁹ Department of Neurology NYU Langone Health New York US

¹⁰ Department of Clinical Neurosciences Division of Neurosurgery University of Cambridge Cambridge GB

¹¹ Myelopathy.org Cambridge GB

¹² School of Surgery Health Education North West Manchester GB

¹³ Department of Neurosurgery Cambridge University Hospitals NHS Foundation Trust Cambridge GB

Corresponding Author:

Benjamin Davies BSc, MBChB, MPhil, PhD

Department of Clinical Neurosciences

Division of Neurosurgery

University of Cambridge

Addenbrooke's Hospital, Hills Road

Cambridge

GB

Abstract

Background: Neurological disorders affect approximately 3 billion people globally, yet clinical trial success is often hindered by poorly chosen outcome measures, impacting trial design, compliance, and interpretation. Core Outcome Sets (COS) have emerged over the past 25 years as standardized tools to enhance outcome selection, ensuring comparability across studies and reflecting the priorities of both researchers and patients. Despite the success of COS initiatives in other fields, their development in neurology remains limited, leaving many trialists without disease-specific guidance. Given common themes across neurological COS, a unified framework—a 'COS of COS'—could support outcome selection where no disease-specific COS exists.

Objective: This study (COS-Neuro) uses Artificial Intelligence (AI) to analyse existing COS, identifying shared outcome domains to develop a thematic framework, streamlining COS creation and improving neurological trial design.

Methods: COS-Neuro was developed using AI-assisted thematic framework analysis, followed by expert review. A modified 6-step thematic analysis was used without pre-determined codes:

1. Dataset Gathering – Data from the COMET database was collected and COS domains for neurological diseases were coded.
2. Prompt Design & Testing – LLMs (ChatGPT 3.5, Google Gemini 1.5 Flash, Meta Llama-2-70b) were trialled, and prompts refined based on responses.
3. Thematic Analysis – LLMs categorised domains into core areas.
4. Human Refinement – Experts reviewed LLM-generated core areas and selected the most appropriate for interpretation.
5. Clinical Validation – Experts validated domains, core areas, and concepts.

This streamlined approach integrated AI with expert oversight to develop a standardised COS framework for neurological disorders.

Results: With the assistance of LLMs, particularly ChatGPT, a robust conceptual framework for COS in neurological disorders was developed based on the existing 112 COS. Adapting OMERACT model, 4 concepts, 13 core areas and 75 domains were finalised following consensus of clinicians.

Conclusions: COS-Neuro establishes AI-assisted recommendations for COS in neurological disorders. This project provides a foundation for future COS research and a reference for trials lacking established COS. It also sets a precedent for AI-assisted qualitative analysis in medicine, with the successful adaptation of OMERACT highlighting its scalability for 'COS of COS' development across specialties.

(JMIR Preprints 23/04/2025:76448)

DOI: <https://doi.org/10.2196/preprints.76448>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <https://www.jmir.org/preprint/76448>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/preprint/76448>

Original Manuscript

Development of a Core Outcome Set for Neurological Disorders (COS-Neuro): an AI-Assisted Thematic Framework Analysis.

Authorship

Dr. Shyun Ping, Tiong, MD. *Portsmouth Hospitals University NHS Trust, Wessex Deanery*; ORCID 0000-0002-8004-397X; dennis_tiong@hotmail.com

Dr. Xiaoyu, Yang, MD, MSc, PhD. *Imperial College London; Imperial College Healthcare NHS Trust*; ORCID 0000-0002-9055-1230; x.yang@nhs.net

Alvaro Yanez, Touzet, BSc (Hons). *University of Manchester*; ORCID 0000-0001-9309-1885

Dr Christopher Paul, Millward, MBBS, PhD. *University of Liverpool*; ORCID 0000-0001-7727-1157; christopherpaulmillward@gmail.com

Dr. Carl M., Zipser, MD, PD Dr. med., FEBN, M.A.; *Spinal Cord Injury Center and Department of Neurology and Neurophysiology, Balgrist University Hospital, University of Zurich, Zurich, Switzerland* ORCID 0000-0002-4396-4796

Dr. Lindsay, Tetreault, MD, PHD; *Department of Neurology, New York University Langone Health, New York, USA*; ORCID 0000-0001-8435-4292

Dr. Ali, Gharooni, MBChB, MSc, MRCS. *Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom*; myelopathy.org, *School of Surgery, Northwest Deanery, UK*; ORCID 0000-0002-6705-1115

Dr. Benjamin, Davies, BSc MBChB (Hons) MPhil PhD. *Department of Neurosurgery, Cambridge University Hospital; Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom*; myelopathy.org ORCID 0000-0003-0591-5069; bd375@cam.ac.uk

Correspondence:

Dr. Benjamin, Davies

bd375@cam.ac.uk

Key words: Core outcome set; Neurological disorder; Artificial intelligence; Large Language Model; Thematic Framework Analysis; ChatGPT



Abstract

Background

Neurological disorders affect approximately 3 billion people globally, yet clinical trial success is often hindered by poorly chosen outcome measures, impacting trial design, compliance, and interpretation. Core Outcome Sets (COS) have emerged over the past 25 years as standardized tools to enhance outcome selection, ensuring comparability across studies and reflecting the priorities of both researchers and patients. Despite the success of COS initiatives in other fields, their development in neurology remains limited, leaving many trialists without disease-specific guidance. Given common themes across neurological COS, a unified framework—a ‘COS of COS’—could support outcome selection where no disease-specific COS exists. This study (COS-Neuro) uses Artificial Intelligence (AI) to analyse existing COS, identifying shared outcome domains to develop a thematic framework, streamlining COS creation and improving neurological trial design.

Methods

COS-Neuro was developed using AI-assisted thematic framework analysis, followed by expert review. A modified 6-step thematic analysis was used without pre-determined codes:

1. Dataset Gathering – Data from the COMET database was collected and COS domains for neurological diseases were coded.
2. Prompt Design & Testing – LLMs (ChatGPT 3.5, Google Gemini 1.5 Flash, Meta Llama-2-70b) were trialled, and prompts refined based on responses.
3. Thematic Analysis – LLMs categorised domains into core areas.
4. Human Refinement – Experts reviewed LLM-generated core areas and selected the most appropriate for interpretation.
5. Clinical Validation – Experts validated domains, core areas, and concepts.

This streamlined approach integrated AI with expert oversight to develop a standardised COS framework for neurological disorders.

Results

With the assistance of LLMs, particularly ChatGPT, a robust conceptual framework for COS in neurological disorders was developed based on the existing 112 COS. Adapting OMERACT model, 4 concepts, 13 core areas and 75 domains were finalised following consensus of clinicians.

Conclusion

COS-Neuro establishes AI-assisted recommendations for COS in neurological disorders. This project provides a foundation for future COS research and a reference for trials lacking established COS. It also sets a precedent for AI-assisted qualitative analysis in medicine, with the successful adaptation of OMERACT highlighting its scalability for 'COS of COS' development across specialties.

Introduction

Neurological disorders are the leading cause of illness and disability worldwide, affecting an estimated 3 billion people (Collaborators, 2024). Translational research is clearly critical to addressing these unmet needs (Turner et al., 2021). Up to March 2025, there were over 10,800 ongoing clinical trials in neurological disorders, accounting for 21.5% total clinical trials registered on clinicaltrials.gov.

Trial success depends on multiple factors, but the choice of trial outcome measure is fundamental. The outcome measure informs the trial size, it can influence compliance and attrition, and ultimately it defines trial interpretation; Put simply, if the outcome measure is poorly chosen, a trial may reach the wrong conclusion and/or struggle to change practice (Andrade, 2015; Davies et al., 2019; KHAN et al., 2020).

One initiative to support outcome selection in clinical trials, that has gained momentum over the last 25 years, are standardised outcome sets also known as Core Outcome Sets (COS) (Tunis et al., 2016). These are a list of critical disease features that should be measured for a given scope, typically a specific disease. Principally designed to better enable research synthesis (by ensuring studies can be compared without bias), their formation through wide stakeholder engagement, including lived experience, can also ensure studies measure a disease comprehensively i.e. do not overlook features that matter most to patients (Williamson et al., 2017). Initiatives such as COMET (Core Outcome Measures in Effectiveness Trials) and OMERACT (Outcome Measures in Rheumatology) have greatly supported their development, including refining the evidence base on how they should be created. A recent impact assessment of the COS for Rheumatoid Arthritis, one of the first examples created, highlights the potential dividends for trialists, with translational success at a field level correlating with increased adoption (now at 82%) (Kordkheili et al., 2024).

The opportunity for COS to support translational science is therefore clear, the challenge remains that their creation and implementation is time consuming and costly. (Liao & Quintana, 2021). So far, only 55 established COS (August 2024) have been developed for neurological disease, including those with multiple COSs on COMET database. Consequently, many neurological trialists do not have a COS as reference, when designing their trial (Millward et al., 2022). Although guidance, such as the OMERACT Filter can be conceptually helpful for trialists, this remains generic for Rheumatology.

Reviewing the Neurological COS there appear many common themes leading us to hypothesize that there may be a common framework to a COS for neurological disease or a 'COS of COS'. Unlike the OMERACT Filter, this sector specific guidance could support trialists where no disease specific COS exists, or COS developers as they develop their disease specific COS.

This study (COS-Neuro) outlines the development of a COS tool for neurological disorders with the assistance of Artificial Intelligence (AI) by establishing the frequency and scope of outcomes previously reported in COS to generate common themes applicable to neurological disorders.

Methods

This is a mixed-methods study, using semi-automated methods with expert review, to propose cross-cutting themes.

Thematic analysis (TA) is a method for analysing qualitative data by developing patterns of themes across a dataset through data familiarisation or coding (Naeem et al., 2023). TA is conceptualised as family of methods for qualitative analysis, which is further sub-categorised into reflexive, coding, framework analysis, thematic content analysis, deductive, inductive, scientifically descriptive TA and artfully interpretative TA.

AI is revolutionary in qualitative research with its cognitive ability in data processing, coding and pattern recognition and has been proven to effectively assist in process of thematic analysis (Christou, 2024). Large Language Models (LLMs) are advanced transformative AI models that have been trained on massive text datasets to comprehend contextual data and the incorporation with Natural Language Processing (NLP) enables them to generate contextually relevant outputs and propose themes by analysing patterns across diverse domains (Gurusamy et al., 2024). With the appropriate prompt, it can be incorporated in research to enhance data visualisation, analysis and summarise scientific text. (Lubiana et al., 2023).

This COS-Neuro project was developed using thematic framework analysis. (Gale et al., 2013). The themes were generated by identifying and interpreting with assistance of LLM prior to review by clinicians with diverse level of experience and expertise.

Several initiatives have been established to help develop and disseminate COS, including Outcome Measures in Rheumatology (OMERACT) and COMET (Core Outcome Measures in Effective Trials) initiative. Amongst the methodology resources created by OMERACT to help the development of COS is the OMERACT Filter, updated in 2019 to Version 2.1. The aim of the filter was to help COS developers ensure content validity, by categorising the full universe of measurable aspects of healthy conditions (Boers et al., 2019). The framework defines a measurement 'Concept', e.g. Pathophysiology vs Impact which maps to a 'Core Area', e.g. Death, Life Impact, Societal/Resource Use, Manifestation and Abnormalities (**Figure 1**).

Building on successes of OMERACT, the key initiative of COS-Neuro is to adapt its filter to develop a COS of available COS in neurological disorders by AI-assisted thematic analysis with expert reviews. A modified 6-step thematic analysis framework was used where the themes were created without any pre-determined codes (Crosley, 2021).

Step 1: Dataset gathering involved gathering data from COMET database and domains coding from COS of different neurological diseases.

Step 2: Prompt design and trial with 3 LLMs (CHATGPT 3.5, GOOGLE GEMINI 1.5 flash, and META Llama-2-70b). The prompts were improved and re-designed based on the response of the LLMs.

Step 3: Thematic analysis with assistance of LLMs to categorise them into their relevant themes which are known as core areas.

Step 4: Output refined with human expertise to compare the core areas analysed by different LLMs from domains and chose best one to proceed for concept interpretation.

Step 5: Clinical validation of domains, core areas and concepts by experts of varying experience.

Step 1- Dataset gathering

COMET database stored all studies relevant to development of COS for use in clinical trials (<https://www.comet-initiative.org/Resources/Database>). The term 'neurology' was used for the search and only published or ongoing studies were included in the dataset during the time of data collection (August 2022). The COS were then identified and extracted for data analysis. The focus is on neurology, neurosurgery and stroke as main specialties.

Table 1 Inclusion and exclusion criteria for selecting existing COS

Inclusion criteria	Exclusion criteria
Neurology-related studies	Unpublished COS
Brain injury	Withdrawn COS
Spinal cord disorders	Ongoing COS
Dementia	
Neurodegenerative diseases	
Brain injury	
Brain tumour	
Functional neurological disorder	

The COS were chosen by reviewers with knowledge and experience for neurological disorders (neurosurgeon and neurologist) with inclusion and exclusion criteria outlined in **Table 1**. Each study was coded based on the COMET ID to facilitate domains cross check by different reviewers. The

established core outcomes for each neurological disorders are longlisted in Microsoft Excel for analysis.

Step 2- Prompt design and trial

Three LLMs (CHATGPT 3.5, Google Gemini 1.5 Flash and Meta Llama 2 70b) were used to analyse the longlist of domains. They were chosen for their easy accessibility and yet highly developed in the world of AI (Buscemi & Proverbio, 2024): ChatGPT builds on the foundation of Transformer Neural Network which enables it to excel in a wide range of Natural Language Processing Task. It predicts the next word in each text sequence based on the precedent context through pre-training. ChatGPT is capable in learning the statistical patterns and structural characteristics of human language and capturing the contextual nuances in the text. Google Gemini is known to outperform human respondents in Massive Multitask Language Understanding (MMLU) tasks, solving grade exams and image recognition by training in multimodal and multilingual datasets. Llama encompasses pre-trained and fine-tuned generative text models ranging from 7 billion to 70 billion parameter (Buscemi & Proverbio, 2024).

Prompt engineering was adapted to design the prompt to execute thematic analysis with LLM. It involved optimising prompt to achieve a desirable outcome with generative model. In this study, instructions + input subgroup was used to construct prompt used for analysis (Amatriain, 2024). The same prompt and longlist of domains were used in all 3 models to maintain consistency.

Role prompting was an effective way to begin the prompt given the task for a specialised topic (Sankar, 2023). Then the prompt was followed by a background information and instructions. The task must be specific with an example to avoid deviation from the expected outcome.

In this project, the prompt design focused on role as medical researcher (a trialists) with specific instructions to perform thematic analysis on core outcome measurements and provide examples of suitable cross cutting themes based on given domains.

Stop criteria for further prompt generation was based on thematic saturation where all the major domains of neurological disorder outcomes were adequately represented in the core set at clinician's consensus and outcomes had been analysed without yielding further insights (Saunders et al., 2018).

The response across the 3 LLMs was then reviewed with feedback for further improvement.

Step 3- Thematic analysis

In-depth inductive thematic analysis was conducted to generate outcomes from the domains. Data were analysed inductively by combination of LLMs and clinical knowledge from clinicians to identify and understand outcomes important for patients (Jones, 2018) and assess the quality of the dataset used in themes generation.

The domains suited for all neurological disorders were identified. With clinician oversight (SPT), responses from each model were compared and commented for further improvement.

Step 4- Refining output with human expertise, and selection of preferred LLM

Each LLMs were prompted to provide a rationale for their response. An example of the prompt “The longlist with domains provided below are extracted from different core outcome sets paper related to neurology. I want the domains to link to the themes with explanations”. With the explanations from generative models, reviewers were able to establish the thought process of LLM.

The output was then compared and the most suitable LLM was chosen by clinicians working on this project to proceed with further analysis.

The criteria for selecting the preferred LLM was:

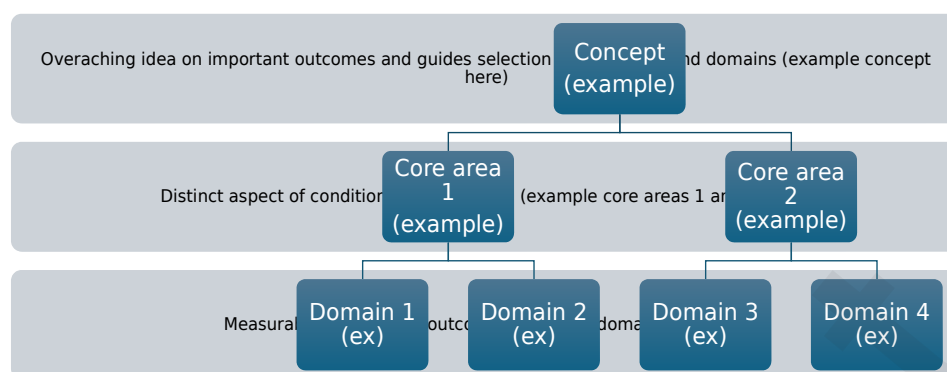
1. Ability to accept prompt long enough for longlist of domains
2. Generate output with relevant domains linked to core areas
3. Core areas deemed acceptable by clinicians

Once the core areas generated by LLM were finalised, a new prompt was designed and tested to link all the relevant domains to the core areas in a table form. It is then extracted to Microsoft Excel.

Step 5- Hierarchical Categorisation using Clinical Expertise, and OMERACT Filter 2.1

Clinicians then worked together, using the OMERACT Filter 2.1 as a reference structure, to aggregate these core areas, into their categories (**Figure 1**). (Boers et al., 2014). This included adapting the descriptive terms used to represent each core area generated by LLM. Experts were encouraged to use their wider medical experience (and not simply rely on the outcomes from existing neurological COS) to help ensure a framework that could generalise to all neurological disorders.

Figure 1: Hierarchical diagram for COS development (adapted from OMERACT Filter 2.1)



Results

Step 1- Dataset gathering

A total of 112 articles on COS relevant to neurological disorders were identified on COMET database and 782 core outcomes were extracted. Each study was coded with the COMET ID and the outcomes are longlisted based on the ID. (**Supplement 1**)

There were 55 neurological disorders with established COS, including those with multiple COS.

Step 2 & 3- Prompt design and initial thematic analysis

The first prompt was designed to instruct LLMs to perform thematic analysis based on the given longlist of domains. LLMs were specifically asked to provide themes with examples for analysis of the thought process.

Prompt 1:

Act as a medical researcher. You are given a longlist of core outcome measurement sets on neurology. Please do an inductive thematic analysis of the longlist and summarise maximum of 20 themes. I want domains as examples that linked to the themes. Example of themes would be comprehensive health assessment with symptoms, blood pressure, mental health as domains.

Longlist of outcomes

Feedback on outcome for prompt 1: ChatGPT and Gemini both focused on symptoms, health status and quality of life which were then selected as the main themes. The length of prompt was too long for Llama. The prompt was further improved by using common themes reported in both ChatGPT and Gemini as examples. Prompt was slightly shortened to accommodate Llama.

The prompt was further edited and improved with feedback on each outcome until a satisfactory output was achieved in prompt 6. Details on the prompts and their outputs are attached in **supplement 2**.

Prompt 6:

Act as a medical researcher. You are given a longlist of core outcome measurement sets in different neurology disorders. please do an inductive thematic analysis of the given longlist and summarise the main themes suitable for COS in neurology. Please use OMERACT as example with concepts, core areas and domains.

Longlist of outcomes

Output:

ChatGPT response

“To conduct an inductive thematic analysis of the given long list of core outcome sets (COS) in different neurology disorders, we will first categorize the provided outcomes into broader themes and then summarize these themes suitable for COS in neurology. As an example, we'll refer to OMERACT (Outcome Measures in Rheumatology) which provides a structured framework for core areas and domains.”

Table 2 ChatGPT generated themes

Themes	Domains
Disease activity/progression	Number of new and/or active lesions
	Number of relapses per year (Annual Relapse Rate)

	Switch from relapsing to progressive
	Cognitive function
	Motor function
	Seizure frequency/severity/duration
Physical Function	Walking ability
	Upper extremity function
	Grip strength
	Balance/Falls
	Fine motor ability
	Ability to perform everyday activities (Activities of Daily Living)
Quality of Life and Patient-reported Outcomes	Health-related quality of life
	Treatment satisfaction
	Patient and family perceptions of care
	Patient mood (Depression, Anxiety)
	Self-esteem
	Social functioning and participation
	Satisfaction with treatment and outcomes
	Overall activity level
	Fatigue
	Pain
	Sleep quality
	Emotional well-being
Neurological Symptoms	Visual functioning
	Bladder and bowel dysfunction
	Sensory impairments (paraesthesia, vibration perception)
	Neuropsychological symptoms
	Psychological symptoms (Anxiety, Depression)
	Cognitive functioning
Health Economics and Health Services	Health economics/cost-utility
	Hospital admissions
	Caregiver burden
	Cost of treatment and care
	Adherence to treatment
Functional Neurological Disorder (FND) Symptoms	Core FND symptoms
	Other physical and psychological symptoms
	Life impact
	Distress and coping mechanisms
Rehabilitation Outcomes	Return to work duration
	Time to return of full function
	Mobility and walking improvements
	Rehabilitation complications and adverse events
	Ability to participate in daily activities
Perioperative and Surgical Outcomes	Operative time
	Postoperative symptom duration
	Complication rate
	Reoperation rate
	Scar sensitivity
	Numbness and sensation in the operating area
	Post-surgery motor function and recovery
Paediatric Neurology Outcomes	Child pain

	Developmental milestones
	Quality of life in children and caregivers
	Pediatric-specific neurological symptoms and disabilities
	School attendance and performance
	Visual and Ocular Outcomes
Visual and Ocular Outcomes	Visual acuity
	Ocular motility and alignment
	Visual field assessment
	Functional vision and reading assessment
	Quality of life related to vision and ocular symptoms
These themes provide a comprehensive overview of the core outcome areas relevant to neurological disorders. COS in neurology should aim to capture these diverse aspects to ensure a holistic assessment of disease impact, treatment efficacy, and patient well-being.	

Step 4- Refining output with human expertise, and selection of preferred LLM

The outputs from ChatGPT, Gemini and Meta were assessed by clinicians based on the criteria listed in the methods. Based on the outcomes, ChatGPT had effectively summarised all the important points which were relevant in neurology. Gemini followed OMERACT pattern and had quite similar concepts to ChatGPT, however the examples were less relevant in neurological disorders. Llama was considered ineffective for this purpose, as it was unable to create a long list of domains or provide reasonable justification. Clinicians therefore agreed to use ChatGPT to proceed with the generated themes in prompt 6 (**Table 2**) as the foundation to further develop COS.

The 10 themes generated by ChatGPT which were then established as foundation for COS-Neuro development were Disease Activity/Progression, Physical Functions, Quality of Life and Patient-reported Outcomes, Neurological Symptoms, Health Economics and Health Services, Functional Neurological Disorder (FND) Symptoms, Rehabilitation Outcomes, Perioperative and Surgical Outcomes, Paediatric Neurology Outcomes and Visual and Ocular Outcomes.

ChatGPT was further prompted to classify and tabulate all the relevant domains into these themes. Two rounds of prompt were used for ChatGPT to successfully filter 782 total domains to 225 domains of relevance to all neurological disorders. It managed to link the domains to the themes/core areas in a table (**supplement 3**).

Prompt 2:

“Please analyse ALL the domains given and classify them into the above themes. Provide your answer in a table with two columns. I want the first column to be ALL the longlist of domains given in the first prompt, second column to be the themes. The themes are disease activity/progression,

physical functions, Quality of Life and Patient-reported Outcomes, Neurological Symptoms, Health Economics and Health Services, Functional Neurological Disorder (FND) Symptoms, Rehabilitation Outcomes, Perioperative and Surgical Outcomes, Pediatric Neurology Outcomes and Visual and Ocular Outcomes”

Step 5 - Hierarchical Categorisation using Clinical Expertise, and OMERACT Filter 2.1

Table 3 outlines the rationale behind each change in the core areas, including rewording, combining, and further sub-categorisation based on clinical appropriation.

The final core areas agreed by clinicians were disease specific activity, Pain, Sensory function, Vision, Autonomic functions, Cognition, Motor functions, Burden of disease and economic impact, Quality of life, ADL, Coping, Social participation, Treatment process, and Child development and formation.

Once the core areas were set, a final round of prompts was designed to aggregate the LLM generated domains, into overarching themes we term main domains. The aim was to create descriptive terms, that would intuitively prompt users of COS-Neuro to understand the construct of each area.

‘Act as medical researcher. You’re developing core outcome measurements in neurological disorders. You’re given the main core areas followed by their domains as text below. I want you to identify the MAIN domains for the core areas.’

225 domains were further narrowed down to 75 domains as tabulated in **Table 4**.

The four main concepts were life impact, neurological function, disease and care manifestations, with a further fifth concept of child development and formation considered optional, depending on the disease in question. Disease and Care Activity is sub-categorising into disease-specific activity, burden of disease and economic impact, treatment process and outcomes to evaluate clinical progression over time and assess effectiveness of therapeutic interventions. Neurological functions consisted of vision, autonomic function, pain, motor, sensory and cognition to establish baseline for patients and understand disease characteristics. Life impact, including quality of life, ADL, coping and social participation covered the social aspects of patients’ life living with neurological disorders. Child development acts an independent core area as a specialised area in neurological disorder. This is considered optional, as it will not be relevant across all conditions. The final framework for COS-

Neuro is illustrated in **Figure 2**.



Table 3: Improvement of core areas with human review

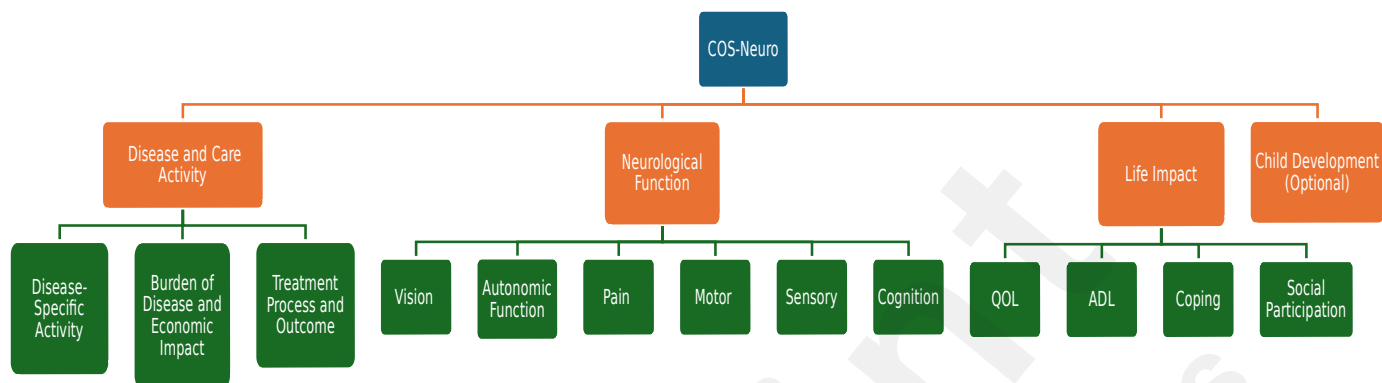
Initial core area	Improved core area	Rationale
Disease activity/progression	Disease-specific activity/progression outcome	The domains in this category focused primarily on seizure activity and control, so disease-specific activity was more appropriate.
Neurological symptoms Visual and ocular outcomes	Pain Sensory function Vision Autonomic functions Cognition Motor functions	Neurological symptoms and visual/ocular outcomes were too broad as core areas, so it was sub-categorising into different neurological functions
Health economics and health service	Burden of disease and economic impact	It was rephrased to healthcare delivery as the domains also consisted of adherence, medication appropriateness, compliance to treatment, education/support which were not limited to economics and services
Quality of life	Quality of life ADL Coping Social participation	QoL had been sub-categorising into QoL, ADL, Coping and social participation as the domains covered a wide range of patient-reported outcomes which were vital in neurological disorders
Rehabilitation outcomes Perioperative and surgical outcomes	Treatment process and outcomes	The two core areas were combined into treatment process and outcomes as they were interconnected and collectively influenced the treatment recovery.
Paediatric neurology outcome	Child development and formation	Paediatric might be easily misunderstood as mixing up with another speciality and the domains such as preschool attendance and performance, engagement in school life were more holistic approach into children's development
Functional neurological symptom disorder	Removed	FND is a diagnosis rather than a core area.

Table 4: AI-identified main domains based on clinicians agreed core areas and concept

Domains	Core area	Concept
Number of new and/or active lesions	Disease-specific activity	Disease and Care Activity
Number of relapses per year (Annual Relapse Rate)		
Switch from relapsing to progressive		
Seizure freedom		
Seizure frequency		
Seizure duration		
Seizure severity		
Hospital admission		
Cost of treatment and healthcare utilization		
Frequency of doctor/clinic visits		
Medication management (adherence, side effects, appropriateness)		
Service provision and related costs		
Hospital admissions and length of stay		
Time to return of full function	Treatment process and outcomes	
Return to work/activity		
Complication and adverse event rates		
Postoperative symptom duration		
Mobility improvement (e.g., walking, standing)		
Surgical outcomes (operative time, incision length, scar)		
Visual field assessment	Vision	Neurological function
Visual functioning		
Eye alignment and movement (ocular motility)		
Functional vision assessment		
Visual fatigue and perception		
Bladder dysfunction and incontinence	Autonomic functions	
Urinary retention		
Faecal incontinence		
Migraine-specific pain	Pain	
Pain due to abnormal sensation		
General pain/discomfort		
Walking and mobility	Motor functions	
Upper extremity function (e.g., grip strength, dexterity)		
Balance and falls		
Muscle strength (e.g., leg, arm)		
Fine and gross motor skills		
Sensation and perception (e.g., paraesthesia, vibratory perception)		
Sensory function in genital and perineal areas		
Neuroma formation		
Cognitive functioning	Cognition	

Memory		
Executive function		
Concentration		
Health-Related Quality of Life (HRQOL)	Quality of life	Life Impact
Global quality of life		
Treatment-related quality of life		
Life satisfaction		
Global health status		
Fatigue		
Mood swings		
Treatment satisfaction		
Activities of daily living assessment		
Self-care		
Occupational role functioning		
Ability to do daily activities		
Alertness		
Sleep quality (total sleep, awakenings, daytime sleepiness)		
Self-management	Coping	
Depression and low mood		
Fear of disease progression or complications		
Patient's ability to understand and respond to communication		
Apathy/indifference		
Patient's self-esteem and self-worth		
Patient and family perceptions of care quality		
Participation in daily functioning	Social participation	
Social support		
Ability to join in activities with others		
Shared decision-making with healthcare professionals		
Social functioning		
Satisfaction with care and services		
Nutrition, growth, and developmental state	Child Development	
Child physical and emotional health		
Child's quality of life		
Academic performance and school attendance		
Developmental milestones		
Caregiver quality of life		

Figure 2 Final framework for COS-Neuro. Four concepts (Orange) were identified; *Disease and Care Activity*, *Neurological Function*, *Life Impact* and *Child Development*. These mapped to 13 Core Areas (Green). QOL Quality of Life. ADL Activities of Daily Living.



Discussion

There is growing awareness and emphasis on using COS in the design and conduct of neurological clinical trials. This is evidenced by their increasing adoption (47% for clinical trials, (Saldanha et al., 2024)), and development (48 COS currently in planning or ongoing (COMET database on 26/03/2024)). That said, many neurological conditions remain unrepresented.

With the assistance of LLMs, particularly ChatGPT, a robust conceptual framework for COS in neurological disorders was developed based on the existing 112 COS available on COMET database. A total of 4 concepts, 13 core areas and 75 domains were finalised following consensus of clinicians. Given the varied presentation of neurological diseases, 75 domains were identified to ensure comprehensive coverage of relevant aspects of disease impact. Functional impairment as a concept could evaluate clinical progression over time and assess effectiveness of therapeutic interventions. Neurological functions establish baseline for patients and understand disease characteristics while patient-centred outcomes covered the social aspects of patients' life living with neurological disorders. Child development and formation acts an independent concept/core area as a specialised area in neurological disorders.

To our knowledge, COS-Neuro is the first project looking into development of 'COS of COS' in neurological disorders and representation of a significant advancement in the field of COS and clinical research by integrating AI into process of thematic analysis. This study by (Lee et al., 2024) has demonstrated ChatGPT's ability to improve the efficiency in thematic analysis with its ability to provide meaningful sets of codes and themes.

Traditional COS development requires manual qualitative data analysis which can be time consuming and labour intensive due to large amount of textual data needed to explore inductively to generate categories and explanations (Pope et al., 2000). The development of Computer Assisted Qualitative Software (CAQDAS) such as MAXQDA, NVivo, ATLAS.Ti or N6 was an important milestone in qualitative research to facilitate data analysis with features such as character-based coding, rich text capabilities and multimedia functions. While CAQDAS is efficient in data transcription and boost accuracy and speed of analysis, it's not widely used by qualitative health researchers (16%) because of the complexity in

mastering them (Zamawe, 2015). These softwares also require paid subscription with MAXQDA costing up to €220 yearly for single user (March 2025). The ability of LLMs to rapidly process large datasets, identify patterns, generate common themes and summarise key points significantly reduces the time, and resources require to establish COS (Naveed, 2023). LLMs are mostly free to use and operate through plain text communication, making them more intuitive than traditional software. However, prompt engineering is pivotal in communication with LLM to ensure relevant, informative and accurate response. A well-crafted prompt can significantly enhance capabilities of AI models (Amatriain, 2024). In this study, prompts were designed and refined with close oversight on outputs by qualified clinicians to optimise the outcomes of common themes.

Advantages and Challenges of AI integration

The integration of AI into thematic analysis offered several advantages:

Efficiency: AI significantly reduced the time required to analyse 782 domains extracted from 112 COS studies, generating 75 meaningful domains relevant to the core areas.

Consistency: ChatGPT demonstrated strong consistency in identifying common themes from longlist of domains with rationale, improving inter-rater reliability.

Scalability: This study establishes a replicable method that extends beyond neurology, providing a framework for AI-assisted COS development in other specialities.

There were 2 notable studies in medical research utilising AI to perform thematic analysis (Kondo et al., 2024; Lee et al., 2024) and both studies suggested that ChatGPT could serve as an assistive tool for thematic analysis to enhance qualitative research efficiency. However, at the time of writing this article, AI-assisted thematic framework analysis has never been used in development of COS to develop relevant themes based on longlist of domains.

AI-integration in thematic analysis presents unique challenges. One major issue is the variability of AI output based on prompt design. Despite using identical dataset and prompt, different LLMs will generate different outcomes due to difference in their underlying training model (Naveed, 2023). Hence, all the three LLMs were given the same prompt to maintain consistency. ChatGPT was chosen for final analysis due to its superior consistency and domain relevance to themes as recognised by clinical expertise.

Researchers also faced with AI ‘hallucinations’ with possibility of inaccurate or misleading information and missing themes or codes. They were addressed by using prompt engineering to refined prompt input to ensure AI-assisted outcomes are reliable and meaningful (Aleksei Turoboc, 2024) and assigning a clinician to lead the engineering and verify its output. LLMs were prompted to provide rationale behind the outcome for clinicians to understand its thought process to enable appropriate refinement. Combining AI-generated outputs with clinicians oversight enable refinement of COS to better reflect real-world clinical priorities (Bragazzi & Garbarino, 2024).

Limitations, Future Development and Implementation

While COS-Neuro represents a significant step towards standardisation of COS in neurological disorders, clinicians have identified several limitations in this study. The datasets used in this study were the outcomes from the existing 112 COS in neurological disorders. The identified core areas were predominantly identified from existing COS, which may not capture complexity of all neurological diseases.

Several important aspects of neurology should be considered in developing COS as recognised by consensus among clinical experts. There seems to be a bias towards demyelinating disease and epilepsy in disease-specific activity core area. Several neurological diseases such as Parkinson’s, Motor Neuron disease, Alzheimer’s are not only relapsing and remitting but rather progressive or stable. Recommendations to include “worsening symptom burden,” “new symptoms,” “accumulation of neurological deficits” as domains could capture wider range of neurological disease. Sleep could be incorporated as well as part of core area with domains including sleep hygiene, sleep latency, insomnia and acting out of dreams. Certain behaviours like hallucinations and delusions could also be reflected as part of core areas given its common presentation in neurological diseases. These domains or core areas were yet to be established as a COS; hence they were not included in the considerations in COS development.

The decision to adapt the OMERACT Filter was pragmatic. Neurological disorders, like rheumatological conditions, are complex and heterogeneous conditions. There are in fact areas of overlap, e.g. neurological manifestations of vasculitis or connective tissue diseases. It recognised that the Filter had been developed for a similar purpose, to improve content validity in outcome selection for Rheumatological clinical trials, and formed via rigorous

development and refinement, exemplified by its use, adoption and evolution over time. The filter is a recommended resource within the COMET Handbook, a step-by-step guide on how to develop a COS. (Kordkheili et al., 2024; Tunis et al., 2017; Williamson et al., 2017).

By leveraging AI-assisted thematic analysis and ongoing advancement of LLMs, future iterations could refine and expand COS-Neuro to include underrepresented neurologic disorders such as functional neurological disorders and rare neurogenetic syndrome.

Nevertheless, COS-Neuro can today serve as a conceptual framework to aid the development of COS for other neurological disorders. It was the authors experience in developing a COS for Degenerative Cervical Myelopathy, that consolidating and categorising a long and diverse range of outcomes was challenging and would certainly have been supported by a conceptual framework like COS-Neuro (Davies et al., 2024). COS Neuro can also assist trialists in selecting appropriate outcome measures while designing trials, particularly for conditions that lack standardised endpoint. Thus, by providing structured framework, it encourages trials measure clinically relevant and patient-centred outcomes, comprehensively, thus improve consistency and comparability across studies (Kirkham & Williamson, 2022; Williamson et al., 2017).

Conclusion

A conceptual framework to inform the selection of outcome measures in neurological trials and/or the development of neurological COS, has been created with assistance from LLM, based on existing neurological COS and the OMERACT Filter 2.1. Of the LLM explored, ChatGPT was considered most effective for this task. The semi-automatic nature of this process can support future iterations of COS-Neuro, and/or application in other diseases, providing a blueprint for 'COS of COS' development.

Acknowledgments

We are grateful to Jeffrey Coates for his assistance with prompt engineering.

We used generative AI tools — ChatGPT (OpenAI, 2023), Gemini 1.5 Flash (GoogleDeepMind, 2024), and LLaMA 2 70B (MetaAI, 2023) — to assist in thematic framework analysis.

Conflicts of interest

The authors declare that they have no conflict of interest.

Abbreviations

Acronym	Definition	
COS-Neuro	Core Outcome Sets in Neurological disorders	An ‘COS of COS’ initiative supporting the development of consensus in outcome measurement for neurological disorders.
COS	Core Outcome Sets	An agreed standardised set of outcomes that should be measured and reported, as a minimum, in all clinical trials in specific areas of health or health care.
OMERACT	Outcome Measures in Rheumatoid Arthritis Clinical Trials	www.omeract.org —An initiative supporting the development of consensus in outcome measurement for arthritis.
COMET	Core Outcome Measures in Effectiveness Trials	https://www.comet-initiative.org - An Initiative in development and application of agreed standardised sets of outcomes (COS)
TA	Thematic Analysis	Research method used to identify and interpret patterns or themes in a data set
AI	Artificial Intelligence	Technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy

LLM	Large Language Models	A type of artificial intelligence algorithm that uses deep learning techniques and massively large data sets to understand, summarize, generate and predict new content
Concept	Overarching idea on important outcomes and guides selection of core areas and domains	
Core area	Distinct aspect of condition or intervention	
Domain	Measurable variables or outcomes from COS	
Outcome	A measurable change or result that can be used to assess the effect of a health care intervention, condition, or disease	
OMERACT Framework	A conceptual framework that helps identify the key areas and domains of health that need to be measured, ensuring a comprehensive approach to outcome measurement	
OMERACT Filter	A tool used to assess and select the appropriate outcome measurement instruments (questionnaires, scales, etc.) for inclusion in the core outcome set.	

References

- Aleksei Turoboc, D. C., Verity Harding. (2024). *Using ChatGPT for thematic analysis*. University of Cambridge. <https://www.bennettinstitute.cam.ac.uk/wp-content/uploads/2024/05/Using-ChatGPT-for-analytics-WP.pdf>
- Amatriain, X. (2024). *PROMPT DESIGN AND ENGINEERING: INTRODUCTION AND ADVANCED METHODS*. arXiv. <https://arxiv.org/pdf/2401.14423>
- Andrade, C. (2015). The primary outcome measure and its importance in clinical trials. *J Clin Psychiatry*, 76(10), e1320-1323. <https://doi.org/10.4088/JCP.15f10377>
- Boers, M., Beaton, D. E., Shea, B. J., Maxwell, L. J., Bartlett, S. J., Bingham, C. O., 3rd, Conaghan, P. G., D'Agostino, M. A., de Wit, M. P., Gossec, L., March, L., Simon, L. S., Singh, J. A., Strand, V., Wells, G. A., & Tugwell, P. (2019). OMERACT Filter 2.1: Elaboration of the Conceptual Framework for Outcome Measurement in Health Intervention Studies. *J Rheumatol*, 46(8), 1021-1027. <https://doi.org/10.3899/jrheum.181096>
- Boers, M., Kirwan, J. R., Wells, G., Beaton, D., Gossec, L., d'Agostino, M. A., Conaghan, P. G., Bingham, C. O., 3rd, Brooks, P., Landewe, R., March, L., Simon, L. S., Singh, J. A., Strand, V., & Tugwell, P. (2014). Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *J Clin Epidemiol*, 67(7), 745-753. <https://doi.org/10.1016/j.jclinepi.2013.11.013>
- Bragazzi, N. L., & Garbarino, S. (2024). Toward Clinical Generative AI: Conceptual Framework. *JMIR AI*, 3, e55957. <https://doi.org/10.2196/55957>
- Buscemi, A., & Proverbio, D. (2024). *ChatGPT vs Gemini vs LLaMA on Multilingual Sentiment Analysis*. ARXIV. <https://arxiv.org/html/2402.01715v1>
- Christou, P. (2024). Thematic Analysis through Artificial Intelligence (AI) *The Qualitative Report* 29(2), 560-576. <https://doi.org/DOI:10.46743/2160-3715/2024.7046>
- Collaborators, G. B. D. N. S. D. (2024). Global, regional, and national burden of disorders affecting the nervous system, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet Neurol*, 23(4), 344-381. [https://doi.org/10.1016/S1474-4422\(24\)00038-3](https://doi.org/10.1016/S1474-4422(24)00038-3)
- Crosley, J. (2021). *What (Exactly) Is Thematic Analysis?* <https://gradcoach.com/what-is-thematic-analysis/>
- Davies, B. M., Khan, D. Z., Mowforth, O. D., McNair, A. G. K., Gronlund, T., Kolias, A. G., Tetreault, L., Starkey, M. L., Sadler, I., Sarewitz, E., Houlton, D., Carter, J., Kalsi-Ryan, S., Aarabi, B., Kwon, B. K., Kurpad, S. N., Harrop, J., Wilson, J. R., Grossman, R.,...Kotter, M. R. N. (2019). RE-CODE DCM (REsearch Objectives and Common Data Elements for Degenerative Cervical Myelopathy): A Consensus Process to Improve

- Research Efficiency in DCM, Through Establishment of a Standardized Dataset for Clinical Research and the Definition of the Research Priorities. *Global Spine J*, 9(1 Suppl), 65S-76S. <https://doi.org/10.1177/2192568219832855>
- Davies, B. M., Yang, X., Khan, D. Z., Mowforth, O. D., Touzet, A. Y., Nouri, A., Harrop, J. S., Aarabi, B., Rahimi-Movaghar, V., Kurpad, S. N., Guest, J. D., Tetreault, L., Kwon, B. K., Boerger, T. F., Rodrigues-Pinto, R., Furlan, J. C., Chen, R., Zipser, C. M., Curt, A.,...Community, A. O. S. R. D. (2024). A minimum data set-Core outcome set, core data elements, and core measurement set-For degenerative cervical myelopathy research (AO Spine RECODE DCM): A consensus study. *PLoS Med*, 21(8), e1004447. <https://doi.org/10.1371/journal.pmed.1004447>
- Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol*, 13, 117. <https://doi.org/10.1186/1471-2288-13-117>
- GoogleDeepMind. (2024). *Gemini 1.5 Flash*. In Google DeepMind. <https://deepmind.google/technologies/gemini>
- Gurusamy, B. M., Rangarajan, P. K., Krishh, P., Keerthinathan, A., Lavanya, G., Meghana, M., Sulthana, S., & Doss, S. (2024). An analysis of large language models: their impact and potential applications. *Knowledge and Information Systems*, 66, 1-24. <https://doi.org/10.1007/s10115-024-02120-8>
- Jones, J. E. (2018). *CORE OUTCOME SET DEVELOPMENT FOR CLINICAL TRIALS: UNDERSTANDING HOW QUALITATIVE RESEARCH APPROACHES CAN HELP TO ACCOMMODATE OUTCOMES THAT ARE IMPORTANT TO PATIENTS* [Doctor of Philosophy thesis]. <https://etheses.bham.ac.uk/id/eprint/9035/1/Jones2019PhD.pdf>
- KHAN, D. Z., DAVIES, B. M., & KOTTER, M. R. N. (2020). Spinal Research — A Field in Need of Standardization. *The Journal of Rheumatology*, 47(4), 633-634. <https://doi.org/10.3899/jrheum.191225>
- Kirkham, J. J., & Williamson, P. (2022). Core outcome sets in medical research. *BMJ Med*, 1(1), e000284. <https://doi.org/10.1136/bmjmed-2022-000284>
- Kondo, T., Miyachi, J., Jonsson, A., & Nishigori, H. (2024). A mixed-methods study comparing human-led and ChatGPT-driven qualitative analysis in medical education research. *Nagoya J Med Sci*, 86(4), 620-644. <https://doi.org/10.18999/nagjms.86.4.620>
- Kordkheili, A. M., Bekker, C., Hoens, A. M., Voshaar, M., Campbell, W., Carmona, L., de Wit, M., March, L., Nielsen, S. M., Shea, B. J., Toupin-April, K., Tugwell, P., Tymms, K., Kelly, A., & Adherence Working, G. (2024). Strategies to promote implementation of core outcomes for medication adherence trials in rheumatology: A report from the OMERACT-Adherence Group. *Semin Arthritis Rheum*, 66, 152411. <https://doi.org/10.1016/j.semarthrit.2024.152411>
- Lee, V. V., van der Lubbe, S. C. C., Goh, L. H., & Valderas, J. M. (2024).

- Harnessing ChatGPT for Thematic Analysis: Are We Ready? *J Med Internet Res*, 26, e54974. <https://doi.org/10.2196/54974>
- Liao, Z., & Quintana, Y. (2021). Challenges to Global Standardization of Outcome Measures. *AMIA Jt Summits Transl Sci Proc*, 2021, 404-409. <https://www.ncbi.nlm.nih.gov/pubmed/34457155>
- Lubiana, T., Lopes, R., Medeiros, P., Silva, J. C., Goncalves, A. N. A., Maracaja-Coutinho, V., & Nakaya, H. I. (2023). Ten quick tips for harnessing the power of ChatGPT in computational biology. *PLoS Comput Biol*, 19(8), e1011319. <https://doi.org/10.1371/journal.pcbi.1011319>
- MetaAI. (2023). *LLaMA 2 (70B parameter model)*. In Meta AI. <https://ai.meta.com/llama>
- Millward, C. P., Armstrong, T. S., Barrington, H., Brodbelt, A. R., Bulbeck, H., Byrne, A., Dirven, L., Gamble, C., Grundy, P. L., Islim, A. I., Javadpour, M., Keshwara, S. M., Krishna, S. T., Mallucci, C. L., Marson, A. G., McDermott, M. W., Meling, T. R., Oliver, K., Pizer, B.,...Jenkinson, M. D. (2022). Opportunities and challenges for the development of "core outcome sets" in neuro-oncology. *Neuro Oncol*, 24(7), 1048-1055. <https://doi.org/10.1093/neuonc/noac062>
- Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research. *International Journal of Qualitative Methods*, 22. <https://doi.org/https://doi.org/10.1177/16094069231205789>
- Naveed, H. K., Asad & Qiu, Shi & Saqib, Muhammad & Anwar, Saeed & Usman, Muhammad & Barnes, Nick & Mian, Ajmal. . (2023). A Comprehensive Overview of Large Language Models. *JOURNAL OF LATEX*. <https://doi.org/10.48550/arXiv.2307.06435>
- OpenAI. (2023). *ChatGPT (GPT-3.5)*. In OpenAI. <https://openai.com/chatgpt>
- Pope, C., Ziebland, S., & Mays, N. (2000). Qualitative research in health care. Analysing qualitative data. *BMJ*, 320(7227), 114-116. <https://doi.org/10.1136/bmj.320.7227.114>
- Saldanha, I. J., Hughes, K. L., Dodd, S., Lasserson, T., Kirkham, J. J., Wu, Y., Lucas, S. W., & Williamson, P. R. (2024). Study found increasing use of core outcome sets in Cochrane systematic reviews and identified facilitators and barriers. *J Clin Epidemiol*, 169, 111277. <https://doi.org/10.1016/j.jclinepi.2024.111277>
- Sankar, S. (2023). *Prompting and prompt engineering? (Part 1)— a comprehensive introduction*. Medium. <https://medium.com/@AlBites/prompting-and-prompt-engineering-a-comprehensive-introduction-29ff6af3211f>
- Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: exploring its conceptualization and operationalization. *Qual Quant*, 52(4), 1893-1907. <https://doi.org/10.1007/s11135-017-0574-8>
- Tunis, S. R., Clarke, M., Gorst, S. L., Gargon, E., Blazeby, J. M., Altman, D. G., & Williamson, P. R. (2016). Improving the relevance and consistency of outcomes in comparative effectiveness research. *J Comp Eff Res*, 5(2), 193-205. <https://doi.org/10.2217/cer-2015-0007>
- Tunis, S. R., Maxwell, L. J., Graham, I. D., Shea, B. J., Beaton, D. E.,

- Bingham, C. O., 3rd, Brooks, P., Conaghan, P. G., D'Agostino, M. A., de Wit, M. P., Gossec, L., March, L. M., Simon, L. S., Singh, J. A., Strand, V., Wells, G. A., & Tugwell, P. (2017). Engaging Stakeholders and Promoting Uptake of OMERACT Core Outcome Instrument Sets. *J Rheumatol*, 44(10), 1551-1559. <https://doi.org/10.3899/jrheum.161273>
- Turner, B. E., Magnani, C. J., Frolov, A., Weeks, B. T., Steinberg, J. R., Huda, N., Shah, L. M., Zuroff, L., Gu, B. J., Rasmussen, H., Edwards, J. G., Save, A. V., Shen, M., Ren, M., Bryant, B. R., Ma, Q., Feng, A. Y., Liang, A. C., & Santini, V. E. (2021). Neurology trial registrations on ClinicalTrials.gov between 2007 and 2018: A cross-sectional analysis of characteristics, early discontinuation, and results reporting. *J Neurol Sci*, 428, 117579. <https://doi.org/10.1016/j.jns.2021.117579>
- Williamson, P. R., Altman, D. G., Bagley, H., Barnes, K. L., Blazeby, J. M., Brookes, S. T., Clarke, M., Gargon, E., Gorst, S., Harman, N., Kirkham, J. J., McNair, A., Prinsen, C. A. C., Schmitt, J., Terwee, C. B., & Young, B. (2017). The COMET Handbook: version 1.0. *Trials*, 18(Suppl 3), 280. <https://doi.org/10.1186/s13063-017-1978-4>
- Zamawe, F. C. (2015). The Implication of Using NVivo Software in Qualitative Data Analysis: Evidence-Based Reflections. *Malawi Med J*, 27(1), 13-15. <https://doi.org/10.4314/mmj.v27i1.4>

Supplementary Files

Untitled.

URL: <http://asset.jmir.pub/assets/b25a6abf28e1b71b37b6b44f072913cc.xlsx>

Untitled.

URL: <http://asset.jmir.pub/assets/96c0edf7abcc3c772dc2ed3f33bed349.xlsx>