

Beatriz Martínez Ruiz

Multi-criteria decision making for ADHD treatment recommendations

MASTER'S THESIS

Supervisors: Dr. Xavier Castells (University of Girona) and Dr. Òscar Raya
(University of Girona)

Tutor: Beatriz López (University of Girona)

Master's Degree in Health Data Science



Tarragona, 2025

Acknowledgements

Transmetre el meu sincer agraïment a totes aquelles persones que han estat al meu costat i m'han donat un cop de mà al llarg d'aquesta etapa.

En primer lloc, als meus supervisors, Xavier Castells i Oscar Raya, i David Ramírez per haver comptat i confiat en mi per la realització d'aquest treball de fi de màster i per la seva orientació, paciència i valuosos comentaris al llarg de tot el procés. El suport ha sigut fonamental per donar forma a aquest treball.

En segon lloc, als meus amics que m'escoltaven i mostraven tot el seu interès al llarg del seu desenvolupament malgrat que no acabessin d'entendre la meitat dels conceptes emprats.

Finalment, vull agrair a la meva família per la seva dedicació. Sense ells i el seu suport incondicional no hauria sigut el mateix. A la meva mare, que ha estat al dia sobre cada avanç del treball i m'ha recordat contínuament que tot l'esforç té la seva recompensa. Al meu pare, per aconsellar-me i recomanar-me en aspectes informàtics. I a la meva germana, per amenitzar-me sobretot les últimes setmanes d'aquesta etapa.

Abstract

Clinical Practices Guidelines (CPGs) suffer from several limitations, such as making general “one-size-fits-all” recommendations. APPRAISE-RS and APPRAISE-RS-E methodologies improve these limitations by incorporating patient and clinician preference and using Multi-Criteria Decision Analysis (MCDA), specifically the Weighted Sum Model (WSM). Both automate, refine, and expand the GRADE heuristic for generating treatment recommendations as used in CPGs. However, both methodologies omit important information, producing suboptimal discriminative and personalized treatment recommendations.

This project aims to deliver a discriminative ranking of personalized treatment recommendations for Attention Deficit Hyperactivity Disorder (ADHD) as a proof of concept. Therefore, several modifications were performed: (i) a more granular benefit/risk-quality scoring system, (ii) expanded weighting ranges, (iii) incorporation of all patient and clinician preferences; (iv) addition of treatment convenience as a new decision criterion. Furthermore, alternative MCDA methods were evaluated to address WSM limitations used in APPRAISE-RS-E.

This recommender system was tested on a dataset of 28 simulated and 65 real patients. Results demonstrated a substantial reduction in tied treatment rankings, indicating improved discriminative power. Additionally, the full incorporation of preferences and the convenience criterion enhanced the personalization of the recommendations. Among the studied MCDA methods, COPRAS (Complex Proportional Assessment) and the modified version of TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) emerged as the most reliable.

The proposed system aligns with principles of patient-centered care and has strong potential for clinical integration. However, further work is needed, particularly in improving treatment cost integration. Future directions include incorporating clinical implementation and expanding to other therapeutic domains.

Keywords: Multi-Criteria Decision-Analysis, treatment recommendations system, evidence based medicine, ADHD, TOPSIS, COPRAS

TABLE OF CONTENT

LIST OF FIGURES	6
LIST OF TABLES	7
LIST OF EQUATIONS	9
GLOSSARY OF ABBREVIATIONS	10
1. INTRODUCTION.....	1
1.1 BACKGROUND WORKS	3
1.2 OBJECTIVES	5
1.3 STRUCTURE AND TIMELINE OF THE PROJECT	6
1.4 SCOPE AND LIMITATIONS	8
2. STATE OF ART	9
3. CONCEPTUAL DESIGN.....	11
3.1 STUDY OF THE MCDA TECHNIQUES.....	12
3.1.1 FINAL MCDA SELECTION	14
3.2 INTRODUCTION OF COST AND CONVENIENCE VARIABLES	14
4. DEVELOPMENT	16
4.1 BENEFIT/RISK-QUALITY SCORES OF INTERVENTIONS FROM APPRAISE-RS AND THEIR WEIGHTINGS IN THE DECISION-MAKING METHOD	17
4.1.1 HANDLING THE COARSENESS OF THE HEURISTIC FOR GENERATING TREATMENT RECOMMENDATIONS	18
4.1.2 ADDRESSING COARSENESS OF THE BENEFIT-RISK JUDGMENTS	19
4.2 PREFERENCES OF PATIENT AND CLINICIAN	21
4.3 APPLICATION OF MCDA TECHNIQUES TO PATIENT AND CLINICAL PREFERENCES AND BENEFIT/RISK-QUALITY ANALYSIS	22
4.3.1 WSM	23
4.3.2 COPRAS	24
4.3.3 TOPSIS	26
4.3.4 MOORA	27
4.4 MERGING PATIENT AND CLINICIAN RANKINGS	28
4.5 INTEGRATION OF FAVOR AND AGAINST RANKINGS	28
4.6 MERGING EFFICACY, SAFETY AND CONVENIENCE RANKINGS	29
5. EXPERIMENTS AND RESULTS	30

5.1	PRELIMINARY EXPERIMENTS USING SIMULATED PATIENT DATA	30
5.1.1	SIMULATED DATASET	31
5.1.2	<i>EVALUATION METRICS OF THE PRELIMINARY EXPERIMENTS</i>	32
5.1.3	<i>RESULTS OF THE PRELIMINARY EXPERIMENTS</i>	33
5.2	EXPERIMENTS WITH THE FINAL STRUCTURE OF THE RECOMMENDER SYSTEM ON THE FINAL VALIDATED DATASET	34
5.2.1	<i>REAL AND SIMULATED PATIENT DATASET</i>	35
5.2.2	<i>EVALUATION METRICS</i>	36
5.2.3	<i>RESULTS OF THIS SECOND ROUND OF EXPERIMENTS</i>	37
5.3	FINAL ANALYSIS	41
6.	DISCUSSION	42
7.	CONCLUSIONS	44
8.	LIMITATIONS AND FURTHER DIRECTIONS	45
9.	REFERENCES	47
10.	ANNEXES	52
10.1	SUPPLEMENTARY TABLES	52

LIST OF FIGURES

Figure 1. GANTT chart of the project development..... 7

Figure 2. Diagram of the project's recommender system 17

Figure 3. COPRAS approach pipeline..... 24

Figure 4. TOPSIS approach diagram 26

Figure 5. MOORA approach pipeline 27

LIST OF TABLES

Table 1. Comparison of the chosen MCDA methods based on the key considerations of the project implementation.....	13
Table 2. GRADE heuristic for generating treatment recommendations from the benefit-risk judgement and its quality of each intervention	18
Table 3. Assigned value to benefit-risk judgment based on evidence quality	19
Table 4. New classification of the effect of intervention based on SMD and 95% confidence intervals	20
Table 5. New classification of the effect of intervention based on RR and 95% confidence intervals	21
Table 6. Summary table of the preferences rating by simulated patients	32
Table 7. Summary table of the preferences rating by simulated clinicians	32
Table 8. Tie distribution across different ranking positions in the preliminary experiments.....	33
Table 9. Number of ties within the top-ranked positions of recommender system on the final dataset when each MCDA method is integrated individually	37
Table 10. Percentage of patients receiving “clear or unequivocal” recommendations within top-x rankings of recommender system on the final dataset when each MCDA method is integrated individually	38
Table 11. Frequency of appearance of pharmacological and psychotherapeutic interventions in top 3 ranking positions across the recommender system on the final dataset when TOPSIS (T2) and COPRAS (C) methods are integrated individually. In the case of TOPSIS, PIS and NIS are manually introduced.	40
Table 12. Number of ties within the top-ranking recommendations of the final recommender system when TOPSIS_2 and COPRAS are integrated individually and APPRAISE-RS-E	41
Table 13. Percentage of patients receiving “clear or unequivocal” recommendations ranking of the new recommender system when TOPSIS_2 and COPRAS are integrated individually and APPRAISE-RS-E	41

Table A1. Percentage of unique recommendations within top-x rankings in the first experiments 52

Table A2. Statistical metrics of the number of recommendations provided by each method in the first experiments..... 52

Table A3. Statistical metrics of the number of recommendations provided by each method in the second experiments..... 52

LIST OF EQUATIONS

Equation 1. Cost of drug in a typical month	14
Equation 2. Decision matrix	23
Equation 3. Weight normalization formula	24
Equation 4. Alternative's score with WSM technique	24
Equation 5. Normalization of the decision matrix	25
Equation 6. Normalized and weighted value of alternative	25
Equations 7 and 8. Maximizing and minimizing indexes, respectively	25
Equation 9. Alternative score with COPRAS technique	25
Equation 10 and Equation 11. n-dimensional Euclidean distance formula	26
Equation 12. Alternative score with TOPSIS technique	27
Equation 13. Min-Max Metric of Tchebycheff formula	27
Equation 14. Calculation of the final score by merging favor and against rankings	39

GLOSSARY OF ABBREVIATIONS

ADHD	Attention Deficit Hyperactivity Disorder
AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
APPRAISE-RS	Automated ParticiPatoRy And personalized trEatment Recommender System
APPRAISE-RS-E	APPRAISE-RS- Expertise
CGI	Clinical Global Impression
COPRAS	Complex Proportional Assessment
CPGs	Clinical Practices Guidelines
EBM	Evidence-Based Medicine
ELECTRE	Elimination and Choice Translating Reality
GRADE	Grading of Recommendations Assessment, Development and Evaluation
IOM	Institute of Medicine
LIIC95%	Lower Interval of the 95% Confidence Interval
LSIC95%	Upper Interval of the 95% Confidence Interval
MCDA	Multi-Criteria Decision Analysis
MOORA	Multi-Objective Optimization by Ratio Analysis
NICE	National Institute for Health and Care Excellence
NIS	Negative Ideal Solution
PIS	Positive Ideal Solution
PROMETHEE	Preference Ranking Organization Method for Enrichment Evaluation
QoL	Quality of Life
RR	Relative Risk
SMD	Standardized Mean Difference
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
VIKOR	Vlšekriterijumsko KOmpromisno Rangiranje
WSM	Weighted Sum Model
WPM	Weighted Product Model

1. INTRODUCTION

Evidence-based medicine and clinical practice guidelines

Making appropriate clinical decisions can often be challenging. In such circumstances, patients, physicians, and other healthcare professionals require reliable and accurate information to support optimal decision-making. This is the fundamental purpose of Evidence-Based Medicine (EBM). EBM integrates the best available research evidence, the clinical expertise of healthcare professionals, and the values and preferences of patients in order to reach the most suitable medical decisions for each individual case. [1, 2]

Rather than promoting a uniform model of care, EBM emphasizes the personalization of clinical decision-making. It does not involve the indiscriminate adoption of findings from recently published literature. Instead, EBM relies on a structured approach that begins with formulating a precise clinical question. This is followed by the critical appraisal and application of the most relevant and high-quality evidence, tailored to address the unique circumstances and preferences of the individual patient. [2]

As medical knowledge is growing uninterrupted, Clinical Practices Guidelines (CPGs) have emerged as a key component of the framework of EBM to bring scientific evidence into clinical practice [3]. These guidelines, crafted by internationally recognized institutions such as GuiaSalud, Agència de Qualitat i Avaluació Sanitàries de Catalunya, the National Institute for Health and Care Excellence (NICE) and the American College of Physicians. CPGs provide systematically developed recommendations designed to assist healthcare professionals in making the best evidence-informed decisions for their patients in specific clinical circumstances. These guidelines synthesize the most relevant and trustworthy research with clinical judgment and patient values, considering the benefits and potential harms of alternative care options to optimize healthcare outcomes. Their main advantage lies in enhancing the quality of patient care by encouraging effective evidence-based interventions while minimizing ineffective or potentially harmful options. [4, 5]

The development and implementation of CPGs have significantly evolved over the past decades, driven by progress in medical research methodologies, digital health technologies, and interdisciplinary collaboration. The Institute of Medicine (IOM) outlines a systematic, stepwise approach [6] beginning with defining clinical questions using the PICO format – Patient, Intervention, Comparison, and Outcome. Once questions are established, relevant studies are identified and reviewed. For each comparison and outcome, meta-analyses may be conducted to synthesize the available evidence. The next steps involve assessing the quality in the evidence generated with meta-analysis (quality in the field of EBM means “confidence in the meta-analysis

results" as these are prone to bias and imprecision), evaluating the balance of benefits and harms for each intervention, and using this analysis to generate specific treatment recommendations. Each recommendation is then assigned a strength, reflecting confidence in its effectiveness and appropriateness. To support this process, the GRADE (Grading of Recommendations Assessment, Development and Evaluation) [7, 8] Working Group has developed a structured methodology for evaluating both the certainty of evidence and the strength of recommendations. GRADE categorizes clinical recommendations into five distinct categories: "strongly recommended," "weakly recommended," "strongly discouraged," "weakly discouraged," and "not rated due to insufficient evidence."

Clinical practice guidelines limitations

Despite the numerous advantages offered by CPGs, such as reducing unwarranted practice variation and improving the translation of research into clinical practice, they are sometimes perceived as inconvenient and time-consuming to implement. This perception often arises from the absence of essential characteristics in some guidelines, such as clarity, specificity, and flexibility [9]. Furthermore, recommendations provided by different organizations on a same topic may differ, which can challenge clinicians to identify the most appropriate course of action for a given clinical situation. [10, 11]

Another major limitation is that recommendations in CPGs can be ambiguous or overly general. Although GRADE aims to offer clarity, its classification into only five levels of recommendation may fall short of capturing the full complexity of clinical decision-making. This is manifested in two ways. Firstly, CPGs often follow a "one-size-fits-all" model that may not account for individual patient differences. While this supports standardization, it can limit their usefulness in real-world settings, especially for complex patients with multiple conditions. Standardized recommendations may overlook important biological, psychological, and social factors, and certain patient subgroups, often underrepresented in research, may not benefit by guideline-based treatments. Secondly, many interventions are recommended at the same level, which can reduce the usefulness of guidelines for clinical decision-making. Without prioritization, clinicians lack clear direction on which treatment to choose first, particularly in time-sensitive or resource-limited situations. This absence of guidance can create uncertainty, especially when selecting the most appropriate option is crucial. [12-14]

Additionally, due to the exponential increase in diagnostic methods and therapeutic options, many CPGs are based on evidence that may already be outdated by the time they are published. According to a study conducted by the Agency for Healthcare Research and Quality, approximately 90% of CPGs remain valid for only 3.6 years [15]. This underscores the need for continuous updates, which demand significant resources and ongoing commitment. [9]

Artificial Intelligence: the solution of CPGs pitfalls

Over the past decades, a significant advancement across multiple fields, including healthcare, has been the development of Artificial Intelligence (AI). AI has introduced notable improvements in diagnosis, treatment, and patient care, with applications ranging from medical image analysis to supporting clinical decision-making and enabling personalized medicine. For instance, AI has enhanced radiology through deep learning-based image interpretation, improving the accuracy in identifying patient with functionally significant coronary artery stenosis [16]. Another example shows the application of AI in radiation oncology to support the decision-making process in terms of treatment planning [17]. More recently, the integration of AI into CPGs is being explored as a means to develop more dynamic, evidence-based recommendations. For instance, AI can help address the issue of guideline obsolescence by automating the systematic review and updating of medical evidence.

Understanding how AI can influence CPGs requires a thorough consideration of both its capabilities and limitations. Ethical, technical, and practical concerns must be addressed to ensure their responsible integration into clinical settings. To build trust and maximize effectiveness, AI systems should be explainable, meaning that their decision-making processes must be clear, interpretable, transparent and understandable to end users. Besides, their recommendations must be grounded in high-quality, reliable data [18].

To address many of the previously mentioned challenges, a research group from the University of Girona developed a methodology known as APPRAISE-RS (Automated Participatory And personalized trEatment Recommender System) [19]. This approach aims to create recommender systems capable of delivering automated, continuously updated, participatory, and personalized treatment recommendations. Subsequently, an enhanced version called APPRAISE-RS-E [20] was introduced, which compares available treatment by weighing selected preference placing greater emphasis on clinicians' preferences over those of patients, acknowledging the clinicians' deeper expertise and clinical knowledge.

1.1 BACKGROUND WORKS

As mentioned, this project is built upon two foundational works: APPRAISE-RS and APPRAISE-RS-E. Both aim to suggest the most appropriate intervention for a given patient by automating, refining, expanding, and iteratively applying the GRADE framework. The APPRAISE-RS is structured into five key phases. [19, 20]

The initial step focuses on collecting essential demographic information from the patient, along with their treatment preferences and those of the clinician. Preferences, in the context of this work, refer to the treatment goals in terms of efficacy and side effects the patient and clinician want to achieve

with the treatment. These preferences are collected using a scale from 1 (least important) to 9 (extremely important). As established by the GRADE framework, only “critical” preferences (those scoring 7 or above) are used. Based on this information, along with relevant data from existing clinical studies, a personalized clinical dataset is generated in the second phase, containing patient characteristics and patient and clinician critical preferences in the second phase. [19]

In the third phase, a set of clinical studies that match patient characteristics is identified. Then, a summary effect is calculated for each intervention and critical preference by meta-analyzing the information available on the selected studies. The quality of the meta-analysis results is then evaluated using a rule-based system, which assigns a confidence level to each result on a four-point scale. By the end of this process, an overall effect of each intervention on each critical preference is obtained along with its quality. [19]

In the next step, a comprehensive benefit-risk assessment is produced by taking into account the combination of the efficacy and safety outcomes for each intervention. Additionally, a quality evaluation of this judgment is carried out. Depending on the risk-benefit relationship and its quality, each intervention is recommended in a ranking with five possible levels: “strongly recommended,” “weakly recommended,” “strongly discouraged,” “weakly discouraged,” and “not rated due to insufficient evidence.” [19]

APPRAISE-RS achieves more updated, personalized and participatory treatment recommendations than those generated in CPGs but still fails to attain fully discriminative treatment recommendations as many of them fall within the same level of recommendation as only 5 predefined levels exist. For this reason, a new version was developed: APPRAISE-RS-E. APPRAISE-RS-E employs the Weighted Sum Model (WSM). It integrates both clinician and patient critical preferences with the benefit-risk data derived from APPRAISE-RS [20]. Although WSM demonstrates solid performance, it has a known limitation: a single criterion with a high weight or scale can disproportionately influence the final score, potentially overshadowing other relevant factors. Furthermore, the WSM implemented in APPRAISE-RS-E does not use available information from non-critical preferences (those scoring 1-6), therefore omitting substantial data.

Despite the advancements of both methods in improving the application of CPGs, they fall short of fully incorporating the preferences of patients and clinicians into the recommendation system. This limitation arises because they are adapted to the GRADE framework, which is primarily designed for human reasoning and lacks fine granularity. As a result, the validity of their recommendations can be limited. In some cases, the system fails to provide a recommendation altogether or presents multiple treatments tied at the same level, which reduces its clinical utility [21].

APPRAISE-RS-E [22] partially achieves the goals sought compared to APPRAISE-RS by ranking recommended treatments, however, it still falls short of the overall objective of the APPRAISE project. It neither personalizes recommendations to the fullest extent nor is as participatory as it could be, because it does not utilize all patient and clinician preferences or incorporate all relevant weightings. Moreover, the lack of granularity means some useful information which could better define the benefit-risk relationship and its quality, and thus the final recommendation, is lost. Furthermore, neither APPRAISE-RS nor APPRAISE-RS-E include important decision-making elements such as treatment convenience (ease or comfort of administration) and costs, which are critical considerations in clinical practice [23].

1.2 OBJECTIVES

The main objective of this project is to deliver a discriminative ranking of more personalized treatment recommendations in the clinical field, with a special focus on Attention Deficit Hyperactivity Disorder (ADHD) as proof-of-concept. This is achieved through the following key approaches:

- Addressing two key limitations of the Cochrane/GRADE heuristic:
 - Considering all preferences from clinicians and patients rather than filtering only the critical ones
 - Integrating the full spectrum of benefit/risk-quality score weightings rather than simplifying them.
- Evaluating various Multi-Criteria Decision Analysis (MCDA) methods.
- Increasing the granularity of the benefit/risk-quality score.

By incorporating input from both healthcare professionals and patients, we will tackle complex clinical decision-making situations that encompass multiple, and sometimes opposing, criteria, resulting in more personalized and discriminative treatment recommendations. In contrast to methods that depend on a single indicator or perspective, MCDA enables the consideration of a wide range of factors, each carrying its own importance, within the overall decision framework.

In this context, the analysis will improve the APPRAISE-RS and APPRAISE-RS-E methodologies to generate more refined recommendation rankings.

Furthermore, several enhancements have been identified and implemented to better support the project's primary aim:

- Incorporate additional decision-making criteria, such as treatment cost and convenience, into the recommender system.

1.3 STRUCTURE AND TIMELINE OF THE PROJECT

The elaboration of this project has spanned approximately five months from February to June 2025; when it is presented in court. The work was organized into four key phases: information research, development, testing, and presentation and report, as shown in Figure 1. To enable a flexible and adaptive workflow, an agile methodology was adopted, allowing for iterative development, regular progress reviews, and continuous integration of feedback throughout the project lifecycle.

Given the current tendency of remote work motivated by the situation of COVID-19 and the geographical differences between universities, it has been developed from home. This has been feasible because the development of the medical software is mainly based on Google Colab, which is an open-source tool for Python programming. Moreover, the project relied on key means including human resources such as project lead, clinical experts and software expert, along with computing equipment.

The information research phase began on February 3, 2025, and focused on familiarizing with the dataset and code base, as well as an acquisition of knowledge about MCDA techniques through a review of available literature. These tasks were completed within the first three weeks and served as the foundation for the subsequent development phase.

The development phase started on February 24, 2025, and involved several iterative tasks. These included adapting the existing code in the Google Colab environment to accommodate new functionalities such as the integration of a new dataset, the implementation of MCDA techniques, and the inclusion of treatment convenience and cost variables. Following each integration, the updated code was evaluated to ensure effectiveness and accuracy, progressing step by step toward the final solution. Each task was scheduled sequentially, allowing for focused improvements and thorough evaluations. This phase concluded successfully in mid-May 2025.

The testing phase included experimentation with the final version of the system and clinical validation using a set of real patient cases. These activities ensured that the system's recommendations were both technically sound and clinically meaningful. This phase was carried out in May 2025.

Finally, the presentation and report phase involved the composition of the written report and preparation for the oral defense. Report writing was carried out over ten weeks beginning April 7, 2025, in parallel with the final development and testing stages. The preparation for the oral presentation was completed during the first two weeks of June 2025.

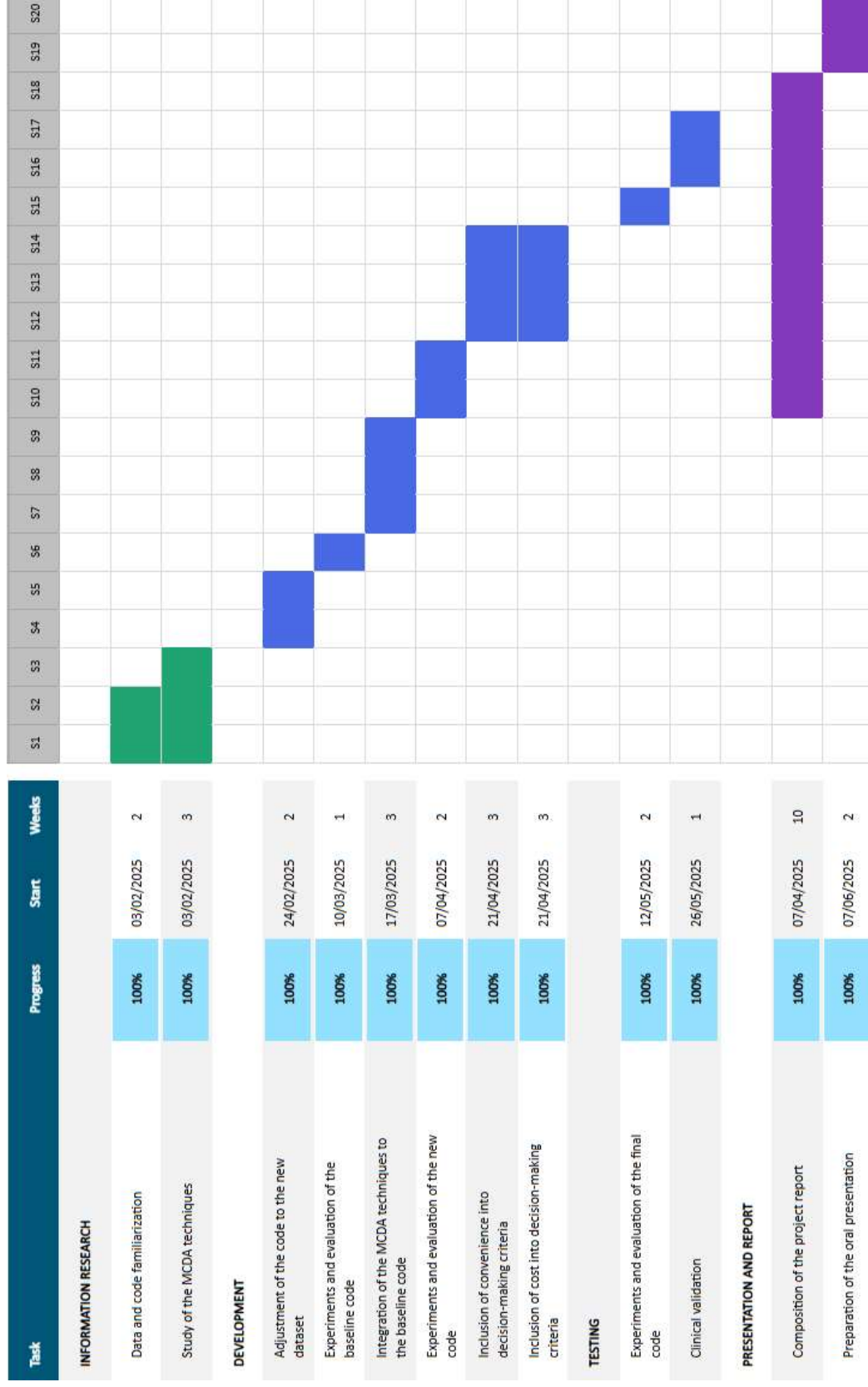


Figure 1. Gantt chart of the project development

1.4 SCOPE AND LIMITATIONS

This project focuses delivering more personalized and discriminative treatment recommendations by analyzing various MCDA methods within the context of ADHD and integrating the best-performing technique, along with minor enhancements, into the APPRAISE-RS and APPRAISE-RS-E frameworks.

The main limitation of the project is the restricted timeframe imposed by the thesis deadlines, set for June 2025. As a result, prospective real-world testing and validation of the final system are not expected to be performed at the end of the timeline, as such that evaluation would require several months of collaboration with clinicians to gather feedback and assess practical utility.

It is important to emphasize that this work is not designed to prescribe treatments directly, but rather to support clinicians in selecting the most appropriate treatment options for each individual patient.

Finally, this project is intended for research and clinical purposes, developed within the academic context of the University of Girona and the Xarxa de Salut Mental I Addicions, the main mental healthcare provider in the province of Girona. At its current stage, the system is a medical decision support system. As such it requires to overcome a regulatory process for its use in the clinical practice [24]. This process takes about 2-3 years and a cost of €1-1.5 million. Thus, its application in clinical or commercial environments lies beyond the current scope. However, the medical device has been turned into an information system by not collecting patient confidential data and providing information on the risk-benefit relationship of treatments upon user request. This information system has recently (22/05/2025) been incorporated into the electronic medical records of the Xarxa de Salut Mental in the province of Girona. The results of this master's thesis will also be incorporated into the information system as soon as the front-end is updated to support the use of the data generated in this work [25].

2. STATE OF ART

A wide variety of MCDA techniques exist to support individuals and organizations in formulating and resolving complex decision problems involving multiple factors. For instance, these methods have been extensively studied in various fields such as environmental contamination, food supply management, and the transition to green energy, among others. However, in medicine, decision-making is even more challenging, as it encompasses not only technical and economic considerations but also deeply human factors, such as individual values, preferences and ethical dilemmas, which can introduce conflicts of interest and complicate the path to a final, shared decision [26].

Despite these challenges, MCDA has gained attention in the healthcare sector, where it has been explored as a tool for supporting a variety of decision-making challenges, including medical device and material selection, disease diagnosis and treatment planning, healthcare operations management, and supply chain logistics [27]. Following the emergence of COVID-19, several research initiatives have applied MCDA techniques, such as TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), AHP (Analytic Hierarchy Process), and VIKOR (Višekriterijumsko KOMpromisno Rangiranje), to automate critical tasks like prioritizing intensive care unit admissions [28] or classifying chest X-ray images using the outputs of multiple convolutional neural networks [29].

Another application of MCDA is in the selection of materials for orthopaedic implants, such as hip [30] or spinal disc replacements [31], to ensure they best meet clinical and biomechanical requirements. The choice of an inappropriate implant material can lead to premature component degradation or complete implant failure, potentially compromising patient safety and requiring costly revision surgeries.

In these studies, MCDA demonstrated the robustness and reliability in material selection by producing stable rankings even under sensitivity analyses and varying criterion weights. In the hip implant case, adjustments of $\pm 15\%$ to key factors like fracture toughness and wear resistance caused only minor shifts in material rankings [30]. Similarly, the spinal disc study showed consistent top-ranked materials across different MCDA methods (fuzzy AHP-TOPSIS, VIKOR and MOORA) [31].

Choosing the most suitable treatment for a patient is a common goal among healthcare professionals. However, this decision is often complex and uncertain, influenced by the clinician's experience and the patient's individual profile and preferences. To assist in such decisions, researchers have developed systems integrating MCDA methods. For example, Elimination and Choice Translating Reality (ELECTRE) III has been applied to support the selection of the

rehabilitation intervention aimed at improving upper limb function in stroke patients [32]. Another study used a hybrid AHP–VIKOR approach to identify the most suitable vaccine for each country by considering both local conditions and vaccine attributes such as production, storage, cost, and potential side effects [33]. In a separate pilot study, an MCDA model was created to better understand and compare the priorities of patients and providers regarding outcomes related to lower-limb amputation levels [34].

Despite the growing interest in applying MCDA to decision-making challenges across various fields, its use in mental health remains relatively underexplored. Diagnosing mental health conditions can be challenging and often requires extended evaluation over multiple sessions. Nonetheless, a few studies have begun exploring MCDA techniques to support clinicians in identifying the most appropriate interventions for individual patients. For instance, one study employed the Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) method to determine the optimal treatment for schizophrenia by weighing factors such as side effects, cost, duration, and efficacy [35]. In another investigation, researchers compared PROMETHEE and TOPSIS to determine the most suitable therapy for managing autism spectrum disorders, tailored to specific patient characteristics [36].

Furthermore, an Indian research team investigated the use of the Multi-Objective Optimization by Ratio Analysis (MOORA) method to pinpoint key psychological factors in patients. This approach aimed to accelerate depression intervention and minimize the risk of worsening symptoms. It was evaluated by performing a correlation analysis between MOORA-identified factors and the Beck Depression Inventory-II model. It revealed a strong agreement, with Pearson's correlation coefficient = 0.8929 and Spearman's rank correlation = 0.7551, indicating consistency in terms of magnitude (how strongly a factor affects depression) and order (which factors are most critical). Moreover, integrating MOORA with conventional therapy significantly enhanced treatment efficiency by reducing session counts by 33-43%, improving detection accuracy to over 82% and accelerating therapy progress by up to 42.93% across depression categories. [37]

3. CONCEPTUAL DESIGN

The initial aim of this thesis was to deliver a discriminative ranking of more personalized treatment recommendations in the clinical field, with a special focus on Attention Deficit Hyperactivity Disorder (ADHD) by integrating an AI-based decision-making technique capable of generating the most distinctive and personalized ranking of available treatments from relevant clinical studies, tailored to a specific patient context. This approach considers both patient clinical data and the clinician and patient preference. In addition, the system is designed to address key ethical concerns, including transparency, explainability, and data privacy.

Decision-making methods use algorithms and models to replicate or enhance human reasoning by processing large volumes of data. These methods can be categorized based on their functionality into rule-based systems, machine learning, deep learning, and analytical decision support systems. For the purposes of this project, some machine learning and deep learning approaches were excluded due to their “black box” nature, which makes it difficult to understand or explain the internal reasoning behind their outputs, an essential requirement in healthcare-related decision support. Examples of such approaches include support vector machines and random forests in machine learning and convolutional neural networks and recurrent neural networks in deep learning.

Rule-based models operate using a series of “if-then” statements derived from expert knowledge to guide decision-making. While these systems are valued for their transparency and interpretability, they can become increasingly complex and unwieldy when required to manage numerous and interdependent criteria, resulting in challenges in terms of scalability and maintenance. In contrast, MCDA techniques stand out for their clarity, flexibility, and structured approach to evaluating alternatives. They enable the construction of decision models that are not only relatively easy to formulate and apply but also maintain a transparent algorithmic foundation. Additionally, MCDA methods are well-suited to incorporating both quantitative and qualitative data, making them particularly useful in healthcare settings where trade-offs between clinical outcomes, patient preferences, and resource constraints are common.

Furthermore, as secondary objective, this project aimed to incorporate additional decision-making criteria beyond efficacy and safety preferences from patient and clinician: cost and convenience.

The following subsections first present a comparison of MCDA techniques, followed by the final selection of the most suitable method. Subsequently, a description of how cost and convenience has been planned to be integrated into the decision-making process is provided.

3.1 STUDY OF THE MCDA TECHNIQUES

MCDA is a structured approach to evaluate and compare different alternatives based on multiple criteria, as its name suggests. The process involves defining the relevant evaluation factors, collecting input from stakeholders regarding their preferences, and constructing a preference model which incorporates these criteria. This model allows for a comprehensive comparison of the alternatives, often resulting in a ranking or classification, and ultimately supports a well-informed decision recommendation.

For this project, we have selected the most studied MCDA methods in the healthcare context: TOPSIS, VIKOR, MOORA, AHP, PROMETHEE, Complex Proportional Assessment (COPRAS) and ELECTRE [27]. Additionally, we included the previously implemented techniques in APPRAISE-RS and APPRAISE-RS-E: the WSM and the Weighted Product Model (WPM). These techniques can be grouped into three main categories based on how they evaluate alternatives: arithmetic, distance-based, and pairwise comparison methods. Following an analysis of their advantages and drawbacks, the chosen techniques for further evaluation in this project were outlined.

Starting with one of the simplest and most well-known methods, **WSM** was originally developed in the 1960s. As its name indicates, it works by calculating the sum of the products of each alternative's evaluation and the weight assigned to each criterion. While it is transparent, flexible and uses straightforward equations that are easy to interpret, WSM may be subject to potential bias as it depends on the linear aggregation of weighted scores. A single heavily weighted or scaled criterion can disproportionately influence the results, potentially overshadowing other relevant factors [38].

Additionally, the method is also highly dependent on the quality of the data used [39]. Incorrect or incomplete data can result in flawed outcomes. For this project, it is essential that weight assignments are determined through a collaborative and agreed-upon process to ensure their accuracy.

A similar technique, known as **WPM**, functions almost identically to WSM but instead relies on multiplicative aggregation, and optionally applies a logarithmic transformation to evaluate each alternative [40]. This introduces a challenge in the project: the input values represent the benefit/risk-quality scores for each alternative and can range from negative to positive, including zero. However, the logarithmic function cannot handle zero or negative values, which can cause computational errors. To resolve this, the data must be adjusted, such as by adding a constant offset or processing negative and positive values separately, to ensure the logarithmic calculations are valid and an error-free.

The last arithmetic-based approach is **COPRAS**, which operates similarly to WSM and WPM but introduces two key distinctions. First, it requires the normalization of alternative values before the method is applied, ensuring comparability across different criteria, hence, only numerical inputs can be used. Second, COPRAS explicitly differentiates between beneficial and non-beneficial criteria during the evaluation process. This allows the method to more accurately reflect trade-offs by maximizing beneficial attributes (e.g., effectiveness) while minimizing non-beneficial ones (e.g., side effects). This dual consideration can lead to more balanced and realistic decision outcomes in complex medical scenarios. [41]

Switching to a different category of MCDA techniques, **TOPSIS** constructs a hypothetical best-case (positive ideal) and worst-case (negative ideal) scenario for each criterion based on the values of the alternatives [42]. It then calculates the Euclidean distance of each alternative from both ideal solutions. In simple terms, this distance represents the length of a straight line between two points in space. The preferred option is the one closest to the positive ideal and farthest from the negative ideal. While the method may initially seem complex due to its mathematical formulation, with proper explanation, clinicians can grasp both the results and the underlying reasoning.

Since TOPSIS is a distance-based method, the computed Euclidean distances are highly sensitive to the input data. Some researchers have observed a phenomenon known as rank reversal, where slight modifications in the values of the alternatives can unexpectedly alter the final ranking of options. This unpredictability is considered one of TOPSIS's significant limitations. [43]

Another distance-based method is **MOORA**, which ranks alternatives using three distinct optimization approaches: the ratio system, the reference point method, and the full multiplicative form [44]. For this project, the reference point method was chosen, as it is considered more objective and interpretable compared to the other two. This method identifies an ideal reference point (typically the best value for each criterion), and alternatives are evaluated based on their deviation from this point. In terms of limitations, MOORA shares similar drawbacks with TOPSIS, particularly its sensitivity to data normalization and the possibility of rank reversal when alternatives are added or removed.

The last method evaluated within the category of ideal-solution-based techniques is **VIKOR**. Unlike the previous approaches, VIKOR is particularly useful when the decision-maker is unable to clearly express preferences among criteria, as it focuses on identifying a compromise solution. This compromise aims to balance two key goals: maximizing group utility and minimizing individual regret, making the final recommendation more acceptable to a group of stakeholders. However, the main drawback of this method lies in its computational complexity. The underlying calculations can be challenging to explain clearly, especially to clinicians without a technical background, potentially hindering their practical usability in healthcare contexts. [45]

The final category of MCDA methods is pairwise comparison, which involves determining the preferred option by comparing alternatives two at a time. The most widely known and applied technique in this category is the **AHP**. AHP simplifies complex decision problems by breaking them down into a structured hierarchy of goals, criteria, sub-criteria, and alternatives. Decision-makers then perform pairwise comparisons at each level of the hierarchy to assess the relative importance of elements, using a scale from 1 (equal importance) to 9 (extreme importance). These comparisons are used to generate priority weights through a consistency-checked matrix, ultimately leading to a ranking of alternatives. [46]

One of AHP's major strengths is that it provides a clear and systematic approach, accommodating both quantitative and qualitative factors. It promotes consistency in judgment and helps make well-justified decisions. However, a notable drawback is its potential complexity and time consumption, especially when the number of criteria and alternatives increases, which can lead to decision fatigue and inconsistency in comparisons.

The next MCDA method considered is ELECTRE, specifically the **ELECTRE II** version, as it provides a final ranking of alternatives rather than simply eliminating weaker options. Moreover, its logic is generally easier to understand compared to other versions of the ELECTRE family [47]. One of the strengths of ELECTRE II lies in its clear and systematic structure, which supports decision-makers by minimizing personal bias and promoting objectivity. Despite this, the method does have some drawbacks. It lacks precision in distinguishing between alternatives that are closely matched, and it does not indicate the degree of superiority between one option and another.

PROMETHEE was the final method selected, given its prior use in two studies focused on treatment selection in mental health. Originally introduced in the early 1980s, PROMETHEE I (partial classification) and PROMETHEE II (complete classification) have gained widespread adoption for tackling multifaceted decision-making scenarios. However, despite their popularity, interpreting the results requires a certain degree of expertise, which may pose a barrier for non-technical users. Additionally, as the number of alternatives and criteria grows, the computational demands and overall complexity can increase significantly. [48]

3.1.1 FINAL MCDA SELECTION

Before selecting the final MCDA methods to be integrated and evaluated in this project, it is essential to define the minimum set of requirements that these methods must meet in order to be compatible with the APPRAISE-RS system and to generate a ranking of alternatives with minimal ties, while effectively showing the benefit-risk balance. These criteria ensure that the selected techniques not only align with the system's architecture and objectives but also uphold standards of transparency, interpretability, and adaptability. Key considerations include:

- Ability to handle both beneficial and non-beneficial criteria
- Produce a clear ranking of alternatives
- Accommodate clinician and patient preferences
- Maintain computational efficiency
- Given the healthcare context, methods must be explainable and user-friendly enough to support shared decision-making between practitioners and patients

A complete and concise comparative table summarizing how each MCDA methods performs across these requirements relevant to the implementation of this project is illustrated in the Table 1.

Therefore, pairwise comparison methods are excluded, as they are not compatible with the APPRAISE-RS methodology for capturing preferences. As described in the introduction section, APPRAISE-RS collects preferences using a direct rating scale from 1 (not at all important) to 9 (extremely important). In contrast, pairwise comparison methods assess the relative importance between pairs of options on a scale from 1 (equal importance) to 9 (extreme importance of one over the other). Converting between these two approaches may lead to the loss of critical information or require significant modifications during the preference-gathering process for both patients and clinicians.

Moreover, these methods rank alternatives separately for each criterion and then aggregate the rankings by applying the corresponding weights. However, the ranking is based solely on the position of each alternative rather than the actual performance values. As a result, two alternatives might end up with the same overall score, even if one has an unsuitable benefit/risk-quality score. This limitation can result in misleading conclusions, especially in sensitive healthcare contexts where precise differentiation is crucial.

On the other hand, the VIKOR technique was not selected for this project due to its complexity, which makes it challenging to explain the ranking generated to end users such as clinicians. The method relies on intricate formulas to rank alternatives, potentially hindering its practical usability. Similarly, WPM was excluded from the final evaluation. Although it operates similarly to COPRAS, it does not account for non-beneficial attributes, such as cost, which are critical and will be incorporated in later stages of the project. Therefore, such factors need to be considered from the outset.

As a result, this project will focus on analyzing the following MCDA techniques: TOPSIS, MOORA, WSM, and COPRAS, as extensions to enhance the recommendation capabilities of the APPRAISE-RS and APPRAISE-RS-E systems.

Table 1. Comparison of the chosen MCDA methods based on the key considerations of the project implementation

Considerations	WSM	WPM	COPRAS	TOPSIS	MOORA	VIKOR	AHP	ELECTRE II	PROMETHEE II
Handling beneficial and non-beneficial criteria	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Production of clear ranking	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Accommodation of clinician and patient preferences	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Computational efficiency	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Explainability	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No
User-friendliness	Yes	Yes	Yes	No	No	No	Yes	No	No

3.2 INTRODUCTION OF COST AND CONVENIENCE VARIABLES

Cost and convenience play a critical role in real-world clinical decision-making. For instance, there is a significant difference between a pharmacological treatment that can be as low as 10€ per month (e.g. methylphenidate immediate release) and a cognitive behavioral therapy that can be as high as 1000€ per month (250€ per session) in some cities. Many families may not be able to afford the most effective option due to financial limitations. Likewise, convenience factors such as the frequency of follow-up sessions or daily pill intake influence treatment adherence. By incorporating these variables, the system better reflects the practical constraints faced by patients and recommends feasible and sustainable interventions.

- **Cost variable**

Cost provides information on the estimated monthly expense of each drug-based intervention. To calculate this, we obtained for each drug the retail price and the number of pills per package and then applied the Equation 1.

$$\text{Monthly cost (€)} = \left(\frac{\text{cost of the package}}{\text{number of pills in the package}} \right) * \text{pills per day} * 30$$

Equation 1. Cost of drug in a typical month

This formula allows us to estimate the total cost a patient would incur over a typical month, assuming consistent daily usage. On the other hand, this calculation does not apply to psychotherapeutic interventions due to the variability in frequency, format (individual or group), delivery method (in-person or telematic), and regional pricing.

The original plan was to use it as a denominator in the final score, applying a Return on Investment logic, which is typically used to assess the efficiency of an investment by comparing the benefits gained to the cost incurred [49]. In the context of this project, the "benefit" would represent the treatment's combined value (efficacy, safety, and convenience), while monthly cost would reflect the "investment".

However, due to insufficient standardized and quantifiable convenience data for psychotherapeutic interventions, apply this logic would lead to incomplete or misleading outputs. Consequently, cost was finally excluded as a scoring factor and is instead presented as a descriptive attribute, providing financial context for clinical interpretation without distorting rankings.

- **Convenience variable**

Convenience, along with efficacy and safety, represents a key dimension of treatment selection. While efficacy ensures the treatment achieves its intended clinical outcomes, safety addresses patient safety and adverse effects. Convenience covers aspects such as ease of administration, accessibility, and adherence-related factors like dosage frequency and formulation.

To account for convenience, two integration strategies were considered: (1) incorporating it as an additional patient and clinician preference, ranked from 1 to 9 within the MCDA, or (2) as an independent variable used to generate a separate ranking, introduced at the end of the recommender system.

After evaluating how much importance should be assigned to convenience, we chose the second option. This decision was driven by the lack of sufficient information on how patients and clinicians would rate this new preference due to the prospective nature of the study and we would only be able to determine its final weight in the decision-making system later.

4. DEVELOPMENT

This section outlines the development of the recommender system, including the following enhancements and modifications made to the APPRAISE-RS-E methodology to align with the objective of the project to deliver more discriminative and personalized treatment recommendations:

1. Considering all preferences from clinicians and patients.
2. Integrating the full spectrum of benefit/risk-quality scores weightings.
3. Evaluating and implementing different MCDA methods.
4. Increasing the granularity of the benefit/risk-quality score.

Additional development related to secondary objectives were:

5. Incorporating additional decision-making criteria, such as treatment cost and convenience, into the recommender system.

The overall structure of the project's recommender system is illustrated in Figure 2.

The first step consists of gathering the summary effects of each intervention on each preference for every patient, as provided by the APPRAISE-RS system. These benefit/risk-quality scores are then translated into numerical scores to facilitate computational processing, and the interventions are then grouped based on this score into two main categories: those in favor and those against the intervention.

In the second step, both patient and clinician preferences are collected and then separated into efficacy and safety categories, resulting in four datasets: patient efficacy preferences, patient safety preferences, clinician efficacy preferences and clinician safety preferences.

The third step is based on applying one of the MCDA previously selected in the literature review to each category, resulting in eight distinct MCDA rankings: efficacy patient in-favor, efficacy patient against, efficacy clinician in-favor, efficacy clinical against, safety patient in-favor, safety patient against, safety clinician in-favor, safety clinical against.

In the fourth step, these rankings, representing both patient and clinician perspectives for efficacy and safety preferences, are combined using the Expertise model, which applies WSM. The predefined weights (70-30) are used to reflect the relative importance of clinician and patient input in the treatment decision-making process.

The fifth step involves integrating the in-favor and against rankings by subtracting the against score from the in-favor score, merging the rankings into a unified output.

In the sixth step, these two rankings were merged with the convenience ranking using a WSM method, assigning 45% weight to efficacy, 45% to safety, and 10% to convenience.

Finally, a ranked list of treatment options is generated, specifically adapted to each patient's context and preferences, as well as clinician preferences.

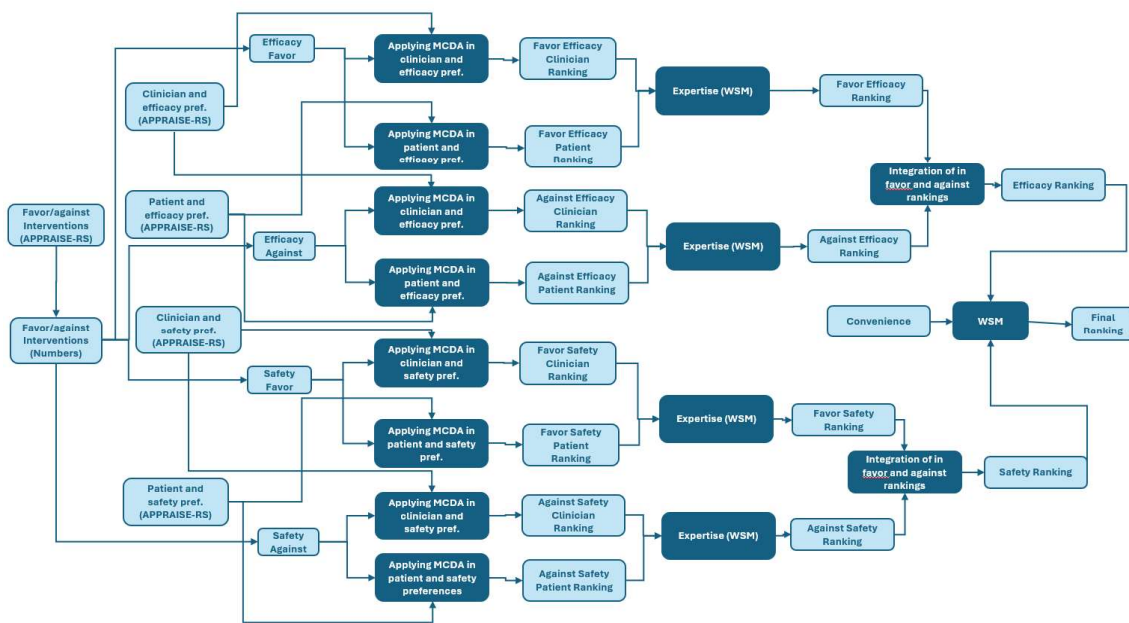


Figure 2. Diagram of the project's recommender system

The remainder of this section details the methodological steps, starting with a deeply explanation of preparation of the input data: benefit/risk-quality scores and patient and clinician preferences. The structure and application of the selected MCDA methods are then described. The final part explains how the nine rankings are merged, first by combining patient and clinician inputs, then favor and against scores, and finally, integrating efficacy, safety and convenience rankings.

4.1 BENEFIT/RISK-QUALITY SCORES OF INTERVENTIONS FROM APPRAISE-RS AND THEIR WEIGHTINGS IN THE DECISION-MAKING METHOD

One of the objectives of this work was to deliver more discriminative treatment recommendations. One reason for this limitation was that the recommender system handled information too coarsely. This coarseness had two sources, namely: both the benefit-risk judgement of the interventions and the heuristic for generating treatment recommendations from the benefit risk analysis and its quality

oversimplified available information. In this work we tackled both, starting with the heuristic for generating treatment recommendations and fishing with the benefit-risk judgement.

4.1.1 HANDLING THE COARSENESS OF THE HEURISTIC FOR GENERATING TREATMENT RECOMMENDATIONS

As described in previous sections, the APPRAISE-RS system evaluates all relevant clinical studies corresponding to a patient's profile. Based on the data reported in these studies, including intervention effects and associated critical preferences, the system aggregates the effects of each intervention on each preference obtained in several studies using meta-analytical methods. This process yields an estimated treatment effect along with a quality assessment reflecting the confidence in the meta-analysis results. By incorporating efficacy and safety outcomes for each intervention, the system determines whether the benefit-risk balance is favorable to placebo, favorable to intervention or indeterminate. By considering the quality of the meta-analyses results for each efficacy and safety variable included in the benefit/risk balance a quality score of this balance is generated [19]. This benefit/risk-quality assessment serves to generate the final recommendation that can fall into one of five categories based on the GRADE heuristic: "strong in favor," "weak in favor," "strong against," "weak against," or "recommendation cannot be made," as shown in Table 2. For example, if the benefit-risk judgment is "favorable" and the quality of the evidence used to determine this judgement is rated as "high" or "moderate", the resulting recommendation is categorized as "strong in favor". [19]

Table 2. GRADE heuristic for generating treatment recommendations from the benefit-risk judgement and its quality of each intervention

		Benefit-risk judgment quality			
		High	Moderate	Low	Very low
Benefit-risk judgment	Favorable to the intervention	strong in favor	strong in favor	weak in favor	weak in favor
	Probably favorable to the intervention	weak in favor	weak in favor	weak in favor	weak in favor
	Neither favorable nor unfavorable	strong against	weak against	weak against	weak against
	Probably favorable to the placebo	weak against	weak against	weak against	weak against
	Favorable to the placebo	strong against	strong against	weak against	weak against

Therefore, a final recommendation is generated for each combination of preference, intervention, and patient profile, and this information is stored in a file. This list of recommendations will serve as one of the key inputs for the system developed in this thesis.

As previously mentioned, clinical recommendations are classified into five categories according to the GRADE guidelines (Table 2). These qualitative categories must be translated into numerical ordinal values (from -2 to +2), as certain MCDA techniques, such as WSM, cannot process textual inputs. Additionally, this transformation enhances interpretability and consistency across each MCDA step.

In the original APPRAISE-RS-E system, strong recommendations were assigned to a value of 2, while weak recommendations were assigned to a value of 1, resulting in a limited scale. As can be seen in Table 2, many situations result in “weak in favor” leading to a loss of information that ultimately yields poorly discriminative treatment recommendations, which we expect to address by increasing input data granularity.

To address this limitation, one of the key improvements introduced in this project was to expand the numerical range of benefit/risk-quality scores, assigning distinct values to both strong and weak recommendations based on the quality of evidence and direction of the benefit-risk judgment.

The final scoring system is presented in Table 3.

Table 3. Assigned value to benefit-risk judgment based on evidence quality

		Benefit-risk judgment quality			
		High	Moderate	Low	Very low
Benefit-risk judgment	Favorable to the intervention	5	4	3	2
	Probably favorable to the intervention	4	3	2	1
	Neither favorable nor unfavorable	-3	-2	-1	-1
	Probably favorable to the placebo	-4	-3	-2	-1
	Favorable to the placebo	-5	-4	-3	-2

This expanded scoring system introduces greater granularity, helping to better reflect differences in the benefit-risk relationship of each intervention, which will ultimately improve clinical recommendations as they will represent more accurately patient and clinician preferences within the MCDA framework leading more discriminative recommendations.

4.1.2 ADDRESSING COARSENESS OF THE BENEFIT-RISK JUDGMENTS

A second source of APPRAISE-RS-E insufficient discrimination in treatment recommendations was that the benefit-risk assessment was too coarse due to the GRADE method being largely analogical. It simplifies the benefit-risk judgment into whether an intervention as favorable or unfavorable in five categories, a format that is straightforward and accessible for clinicians to interpret but failing to introduce nuances in such judgement.

To address this, we developed a second, more granular approach that expands the benefit-risk judgment from 5 to 9 categories based on the guidance provided in the Cochrane Handbook [50], while maintaining the 4-point scale for evidence quality. This enhanced structure is designed to capture more variation between interventions, thereby improving the discriminatory power of the system.

To understand how efficacy and safety analysis is performed the way effect size of efficacy and safety variables must be introduced. The judgments on the efficacy/safety of interventions are based on the effect size on the interventions on efficacy/safety outcomes (preferences), which are expressed as standardized mean differences (SMD) and relative risks (RR) for continuous and binary variables, respectively. In this point it is worth stressing that these SMD and RR were obtained by applying meta-analysis techniques as explained in previous sections.

Table 4 and 5 outline the how we handled efficacy and safety effects (SMD and RR) to generate judgments on the efficacy and safety of the interventions, which are lately used to generate a judgment on the risk-benefit balance (using rules: APPRAISE-RS and WSM: APPRAISE-RS-E). In this point of the work, defined 9 levels of judgments on the efficacy and safety, as opposed to GRADE's 5. In table 4, the classification system based on SMD ranges and the 95% Confidence Interval lower and upper bounds (LIIC95%, LSIC95%) is reported [51]. Incorporating the lower limit LIIC95% ensures that the estimated effect is not only statistically significant but also clinically meaningful. A positive LIIC95% indicates a high level of confidence in a beneficial effect, while a negative LIIC95% suggests potential harm or lack of benefit. On the other hand, the SMD values follow Cohen's thresholds [52]: 0.2 (small), 0.5 (moderate), and 0.8 (large) effects. Finally, each judgment category (J1-J9) corresponds to a specific score, depending on the quality of the supporting evidence (ranging from 4 to 1).

Table 4. New classification of the effect of intervention based on SMD and 95% confidence intervals

Conditions	Judgment	Benefit-risk judgment quality			
		4	3	2	1
$SMD \geq 0,80$ AND $LIIC95\% > 0,00$	J1	+5	+5	+4	+4
$0,50 \leq SMD < 0,80$ AND $LIIC95\% > 0,00$	J2	+4	+4	+3	+3
$0,20 \leq SMD < 0,50$ AND $LIIC95\% > 0,00$	J3	+3	+3	+2	+2
$0,00 < SMD < 0,20$ AND $LIIC95\% > 0,00$	J4	+2	+2	+1	+1
$LIIC95\% \leq 0,00$ AND $LSIC95\% \geq 0,00$	J5	0	0	0	0
$-0,20 < SMD < 0,00$ AND $LSIC95\% < 0,00$	J6	-2	-2	-1	-1
$-0,50 < SMD \leq -0,20$ AND $LSIC95\% < 0,00$	J7	-3	-3	-2	-2
$-0,80 < SMD \leq -0,50$ AND $LSIC95\% < 0,00$	J8	-4	-4	-3	-3
$SMD \leq -0,80$ AND $LSIC95\% < 0,00$	J9	-5	-5	-4	-4

The scoring system can also be based on the RR ranges and the 95% confidence intervals [51], as illustrated in Table 5. Each RR interval is categorized into judgment levels J1 through J9, representing a gradient from highly favorable to highly unfavorable treatment effects. The closer the RR is to zero (with $LSIC95\% < 1$), the stronger the benefit. As RR increases beyond 1 (with $LSIC95\% > 1$), the interpretation shifts toward potential harm. Including $LIIC95\%$ and $LSIC95\%$ helps determine whether the estimated effect is consistently beneficial across the confidence range. When $LSIC95\% < 1$, the entire confidence interval supports a benefit ($RR < 1$), increasing confidence in the intervention. Conversely, $LIIC95\% > 1$ suggests the intervention may increase harm ($RR > 1$), and this is reflected in increasingly negative scores as the RR grows.

Table 5. New classification of the effect of intervention based on RR and 95% confidence intervals

Conditions	Judgment	Benefit-risk judgment quality			
		4	3	2	1
$RR \leq 0,33$ AND $LSIC95\% < 1,00$	J1	+5	+5	+4	+4
$0,33 < RR \leq 0,54$ AND $LSIC95\% < 1,00$	J2	+4	+4	+3	+3
$0,54 < RR \leq 0,82$ AND $LSIC95\% < 1,00$	J3	+3	+3	+2	+2
$0,82 < RR \leq 1,00$ AND $LSIC95\% < 1,00$	J4	+2	+2	+1	+1
$LIIC95\% \leq 1,00$ AND $LSIC95\% \geq 1,00$	J5	0	0	0	0
$1,00 < RR \leq 1,22$ AND $LIIC95\% > 1,00$	J6	-2	-2	-1	-1
$1,22 < RR \leq 1,86$ AND $LIIC95\% > 1,00$	J7	-3	-3	-2	-2
$1,86 < RR \leq 3,00$ AND $LIIC95\% > 1,00$	J8	-4	-4	-3	-3
$RR \geq 3,00$ AND $LIIC95\% > 1,00$	J9	-5	-5	-4	-4

These new scoring schemes were introduced during the second round of experiments, following the generation and expert validation of the final dataset. Its purpose was to enhance the granularity of clinical recommendations and improve the system's ability to differentiate between interventions more effectively.

Additionally, any intervention without supporting clinical evidence, considering the specific combination of preference, drug, and patient profile, was assigned a score of -1, serving as a penalization factor. This value was chosen instead of 0 to ensure that interventions with insufficient evidence do not rank higher than those with benefit/risk-quality score of 0, thereby reinforcing evidence-based prioritization. In the case that an intervention had at least one preference supported by evidence, the remaining preferences without evidence were marked as 'No evidence'.

4.2 PREFERENCES OF PATIENT AND CLINICIAN

In parallel with the previous subsection, the granularity of the way clinician and patient preferences were handled was expanded. This change was introduced to achieve two objectives: to generate more discriminative and personalized recommendations. By capturing a broader and more nuance range of preferences weightings, the system introduces greater variability between patient profiles,

thereby increasing the discriminative power of the recommendations. Additionally, this refinement supports more personalized outcomes by integrating a wider spectrum of individual patient needs and values.

The origin of the insufficient granularity of preferences arises from the GRADE heuristic. Both patient and clinician preferences were rated on a 1-to-9 scale. This evaluation adheres to the GRADE methodology, which categorizes outcomes into three significance levels [53]:

- Scores 7–9: Outcomes deemed critical for making decisions.
- Scores 4–6: Outcomes considered important but not essential for decision-making.
- Scores 1–3: Outcomes regarded as low importance; these are typically excluded from evidence summaries and have little impact on recommendation development

In its original implementation, APPRAISE-RS-E only considered preferences rated as critical, meaning that their score is higher or equal to 7. While this aligns with GRADE's emphasis on prioritizing high-importance outcomes, it also introduces a significant limitation, namely, the exclusion of moderate and less relevant preferences. This can lead to the loss of valuable insights into patient values and clinician expertise.

To overcome this limitation, the current project expands upon the original approach by deleting this filter and incorporating all preferences, regardless of their assigned score.

Moreover, efficacy preferences were separated from safety preferences. This design ensures that both components receive equal weight in the decision-making process, preventing one from disproportionately influencing the final, especially since the number of efficacy and safety preferences differ. Moreover, generating objective results for each preference type independently may offer greater clarity and interpretability for clinicians, the system's primary end-users, by providing a more transparent breakdown of the trade-offs between treatment efficacy and safety.

4.3 APPLICATION OF MCDA TECHNIQUES TO PATIENT AND CLINICAL PREFERENCES AND BENEFIT/RISK-QUALITY ANALYSIS

Another limitation of the APPRAISE-RS-E lies in its use of the WSM in which a single preference with a significant weight can overly dominate the aggregate score. This dominance can overshadow other relevant clinical or patient-centered factors, ultimately leading to less discriminative treatment rankings. Therefore, alternative MCDA can create more discriminative recommendations.

After a comprehensive theoretical review of the most well-studied MCDA methods, four techniques were ultimately selected based on the specific requirements of this project: TOPSIS, MOORA, WSM, and COPRAS. Each of these MCDA approaches is built upon four fundamental components:

- **Alternatives:** These are the different options considered during the decision-making process. In this context, the alternatives refer to medical interventions, including both pharmacological treatments and non-pharmacological (psychological, neurorehabilitation, and others) approaches.
- **Criteria:** These are the attributes or factors used to evaluate the alternatives. They can be quantitative, such as convenience, or qualitative, such as patient and clinician preferences.
- **Weights:** Assigned values to each criterion based on its perceived importance. They help emphasize more relevant factors over less critical ones in the decision process.
- **Evaluation:** This involves assessing each alternative according to the defined criteria. The resulting scores are then combined with the criterion weights to generate an overall ranking of the alternatives. In this project, this parameter corresponds to the benefit/risk-quality score.

These MCDA techniques are implemented after all input data has been fully processed and organized. In this project, data preparation involves collecting preference ratings from both patients and clinicians and converting the clinical recommendations into numerical ordinal values, named weights $[w_1, w_2, \dots, w_n]$. Based on these preferences, two separate decision matrices are created, one reflecting patient input and the other reflecting clinician input, as illustrated in Equation 2.

$$X = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}; i = 1, \dots, m \quad j = 1, \dots, n$$

Equation 2. Decision matrix

According to the matrix of Equation 2, a_{ij} is the element of the decision matrix for alternative i and criteria j .

The following sections will provide a detailed explanation of how each method was implemented and their respective characteristics.

4.3.1 WSM

To begin with the most straightforward and widely recognized MCDA method, WSM operates effectively only when all input data shares a common unit of measurement. In this particular case, this requirement is satisfied, as all preferences are assessed using a uniform rating scale. Another key aspect of WSM is that it requires the weights assigned to each criterion to be on a normalized scale, that is, the weights must collectively add up to 1. This normalization ensures the criteria are directly comparable, enabling a meaningful aggregation of scores across alternatives.

To normalize the weights, a standard approach is followed: each individual weight (w'_j) is divided by the total sum of all weights, Equation 3. This ensures that every criterion's influence is proportionally adjusted within the overall decision framework.

$$w_j = \frac{w'_j}{\sum_{j=1}^n w'_j}$$

Equation 3. Weight normalization formula

Let's define w_j as the normalized weight associated with the criterion j and a_{ij} as the performance score of alternative i with respect to j then, the overall score of alternative i in WSM, denoted as $A_i^{wsm-score}$, is calculated using the Equation 4.

$$A_i^{wsm-score} = \sum_{j=1}^n w_j a_{ij}, \text{ for } i = 1, 2, 3, \dots, m$$

Equation 4. Alternative's score with WSM technique

Finally, the best alternative is the one that achieves the highest overall performance score.

4.3.2 COPRAS

Continuing with another arithmetic method, COPRAS technique defines a direct and proportional link between the importance of alternatives and the criteria used to evaluate them, incorporating both the assigned weights and performance values that characterize each alternative. This approach selects the optimal decision by taking into account both the most favorable and least favorable scenarios. The pipeline followed in this section is presented in Figure 3.



Figure 3. COPRAS approach pipeline

The first step in this pipeline involves normalizing the assigned weights using the same procedure applied in the WSM method. However, unlike WSM, the COPRAS technique also requires the normalization of the decision matrix. For this, normalization is carried out according to the method outlined in Equation 5.

$$a_{ij}^{norm} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^m a_{ij}^2}}; j = 1, \dots, n$$

Equation 5. Normalization of the decision matrix

Where a_{ij}^{norm} is the normalized a_{ij} weight, i indexes the alternatives and j indexes the criteria.

Then, each normalized value within the decision matrix is multiplied by the weight corresponding to its respective criterion, as shown in Equation 6.

$$a_{ij}^{*norm} = a_{ij}^{norm} w_j; i = 1, \dots, m j = 1, \dots, n$$

Equation 6. Normalized and weighted value of alternative

Where a_{ij}^{*norm} is the weighted normalized value, a_{ij}^{norm} is the normalized value from Equation 4, w_j weight assigned to criteria j and, i and j index the alternative and criteria, respectively.

The next step involves calculating the maximizing and minimizing indices (S_{+i} and S_{-i} , respectively) for each preference, depending on whether they represent beneficial or non-beneficial criteria, using Equations 7 and 8, respectively.

$$S_{+i} = \sum_{j=1}^p a_{ij}^{*norm}; \quad S_{-i} = \sum_{i=p+1}^n a_{ij}^{*norm}; \quad i = 1, \dots, m$$

Equations 7 and 8. Maximizing and minimizing indexes, respectively

The final score for each alternative ($A_i^{copras-score}$) is calculated using the Equation 9.

$$A_i^{copras-score} = S_{+i} + \frac{\sum_{i=1}^m S_{-i}}{S_{-i} \sum_{i=1}^m \frac{1}{S_{-i}}}$$

Equation 9. Alternative score with COPRAS technique

Finally, the interventions are ranked in descending order according to their final scores, with the highest-scoring option receiving the top rank.

4.3.3 TOPSIS

Unlike the previously discussed MCDA methods, TOPSIS is a distance-based approach. It works based on a simple and quite logical principle: the optimal alternative is the one that lies closest to the hypothetical ideal solution and furthest from the hypothetical negative (or worst-case) solution in a multi-dimensional feature space. An important advantage of TOPSIS is its ability to handle both benefit and cost-type criteria effectively. The implementation of this method in the project follows the workflow outlined in the Figure 4.



Figure 4. TOPSIS approach diagram

The initial two steps of the TOPSIS process focus on preparing the method's two core inputs: criterion weights and alternative scores. Both datasets undergo normalization using Equations 3 and 5, respectively. After normalization, each value in the decision matrix is multiplied by its corresponding normalized weight.

Next, the Positive Ideal Solution (PIS) and Negative Ideal Solution (NIS) are determined. In the context of TOPSIS, these are hypothetical alternatives constructed by selecting the best and worst values across all criteria. Specifically, the PIS consists of the maximum values for benefit criteria and the minimum values for cost criteria. Conversely, the NIS contains the minimum values for benefit criteria and the maximum for non-beneficial criteria.

To rank the alternatives, TOPSIS uses Euclidean distance to calculate how far each option is from both the PIS and NIS.

$$S_{i+} = \sqrt{\sum_{j=1}^n (a_{ij}^{*norm} - a_j^{PIS})^2} \quad S_{i-} = \sqrt{\sum_{j=1}^n (a_{ij}^{*norm} - a_j^{NIS})^2}$$

Equation 10 and Equation 11. n-dimensional Euclidean distance formula

Where S_{i+} and S_{i-} are the Euclidean distance of alternative i to the PIS and NIS, respectively, a_{ij}^{*norm} is the weighted normalized value, a_j^{PIS} is the PIS for criteria j and a_j^{NIS} is the NIS for criteria j .

Then, the overall score ($A^{topsis-score}$) for each alternative is calculated using the Equation 12.

$$A^{topsis-score} = \frac{S_{i-}}{S_{i+} + S_{i-}}, \text{ for } i = 1, \dots, m$$

Equation 12. Alternative score with TOPSIS technique

Finally, the best alternative is the one with the highest overall score, meaning it is closest to the PIS and farthest from the NIS, thereby reflecting the most balanced combination of attributes.

4.3.4 MOORA

The final MCDA technique evaluated in this project is MOORA. Its primary objective is to rank a set of alternatives by simultaneously optimizing multiple, often conflicting, objectives while possibly considering constraints. While it shares some procedural similarities with TOPSIS, particularly in requiring normalization of both weights and alternative scores (as shown in Equations 3 and 5), its core mechanism differs substantially.



Figure 5. MOORA approach pipeline

Unlike TOPSIS, which relies on computing the distance to both a PIS and NIS, MOORA's reference point approach focuses on a single ideal point. For each criterion, this reference value is determined based on the type of criterion: for maximization (beneficial) criteria, it is the maximum value among all alternatives, and for minimization (non-beneficial) criteria, it is the minimum value.

The subsequent step involves computing the distance of each alternative from this reference point, essentially measuring how closely an alternative approximates the ideal. In this context, the maximal objective reference point is considered more realistic and less subjective, as the reference coordinates (a_i) are not arbitrarily chosen but are instead values that actually exist within one of the evaluated alternatives. This enhances the interpretability and practicality of the method when applied to real-world decision-making scenarios, such as treatment selection.

In this approach, the performance index ($A_i^{moora-score}$) represents the total deviation of the alternative i from the reference point, taking into account both beneficial and non-beneficial criteria. This index can be mathematically expressed as Equation 13.

$$A_i^{moora-score} = \min_j \{ \max_i |a_{ij}^{*norm} - a_i| \}$$

Equation 13. Min-Max Metric of Tchebycheff formula

The most favorable alternative is the one that exhibits the lowest overall deviation from the reference point, taking into account both beneficial and non-beneficial criteria. In other words, the optimal choice is the alternative with the smallest performance index.

4.4 MERGING PATIENT AND CLINICIAN RANKINGS

Once the MCDA techniques are applied separately to the groups of recommendations in favor and against, derived respectively from patient and clinician preferences, each subgroup yields its own ranking of interventions. These four resulting rankings (favor-patient, favor-clinician, against-patient, and against-clinician) represent different perspectives on the suitability of each intervention.

To derive a comprehensive and balanced recommendation, the rankings need to be integrated. The first step involves combining the patient and clinician rankings within each group (favor and against) using aggregation strategy such as Expertise model. This produces a unified "favor" ranking and a unified "against" ranking.

The Expertise model is an aggregation technique based on WSM, where the final ranking of interventions is determined by weighting the input from both the patient and the clinician according to their relative importance in the decision-making process. In this approach, a weight of 0.30 is assigned to the patient's input and 0.70 to the clinician's input, reflecting the balance recommended by the APPRAISE-RS-E system, which emphasizes clinical expertise while still considering patient values significantly.

Each intervention is assessed using the final scores obtained from the individual MCDA results of the patient and clinician and their respective weights to produce a composite score for each intervention. Finally, interventions are ranked in descending order based on these composite scores, with the highest-scoring alternative considered the most preferred.

4.5 INTEGRATION OF FAVOR AND AGAINST RANKINGS

The next integration step consists of merging these two consolidated rankings, favor and against, into a single, coherent list of prioritized interventions. This ensures that the final output reflects a logical balance between supportive and opposing assessments, while respecting the diversity in preferences and evidence. The integration is performed by subtracting the score of the 'against' ranking from the score of the 'favor' ranking for each intervention. This differential score highlights the net benefit, allowing interventions with stronger supporting evidence and fewer drawbacks to be ranked higher, while still accounting for the presence of negative factors.

4.6 MERGING EFFICACY, SAFETY AND CONVENIENCE RANKINGS

As mentioned, we generated a convenience ranking based on treatment characteristics. This ranking is then merged with the previous rankings for efficacy and safety using a WSM, assigning weights of 45% to efficacy, 45% to safety and 10% to convenience.

The outcome is a refined ranking of interventions that considers both the level of clinical recommendation and the importance of individual preferences, offering a more personalized decision-making support system for clinical contexts.

5. EXPERIMENTS AND RESULTS

During the development of the final project solution, multiple experiments were conducted to validate each step of the process. Each experiment is accompanied by a description of the dataset employed at that stage and a specification of the evaluation metrics applied.

The first experiments focused on evaluating the initial modifications and enhancements made to the APPRAISE-RS-E methodology, as detailed in the first subsection. These trials were conducted while the final dataset was still being compiled and clinically validated by experts, with the objective of determining whether the proposed adjustments aligned with the project's main goal. Additionally, a preliminary analysis of the selected MCDA methods was performed to obtain an initial understanding of their performance with the available data.

On the other hand, the second round of experiments aimed to test the recommender system using the final configuration of the fully validated dataset, in order to extract insights of its performance from a clinical and technical perspective. Furthermore, the treatment ranking generated by the system, particularly those using the best-performing MCDA methods, were analyzed to assess the overall quality of the recommendations.

Finally, the concluding experiment was carried out to compare the performance of the newly developed recommender systems, integrating of the top-performing MCDA methods with original APPRAISE-RS-E methodology.

The following section provides a detailed explanation of all experiments and their corresponding results.

5.1 PRELIMINARY EXPERIMENTS USING SIMULATED PATIENT DATA

With the aim of moving closer to achieving the main objective of this project, several initial modifications were applied on the APPRAISE-RS-E framework during the preliminary assessments. On one hand, to enhance the personalization of recommendations, the full spectrum of patient and clinician preferences was incorporated. On the other hand, to improve the discriminative power of the rankings, the granularity of benefit/risk-quality score of intervention and their corresponding weightings in the decision-making process was increased and alternative MCDA methods were explored in place of the original WSM approach.

While the definitive real-patient dataset was still under development and undergoing clinical validation, the enhanced recommender system was initially evaluated using a simulated dataset constructed with input from two child psychiatrists and two adult psychiatrists. This dataset comprises 28 simulated individuals representing a wide range of ADHD profiles, from typical to

more complex cases, thereby enabling preliminary assessment of how the new MCDA approaches and improvements affect the ranking outcomes.

5.1.1 SIMULATED DATASET

- **Patient characteristics**

The analysis of key demographic variables of the simulated dataset reveals a higher representation of male patients ($n = 20$) compared to female patients ($n = 8$). Regarding age, the dataset spans from 7 to 53 years old, with most patients falling into the adolescent group (11 to 18 years). The remaining individuals are distributed across childhood (under 11), young adulthood (19–35), and middle adulthood (36–53), providing a diverse range of ADHD profiles for evaluation.

In addition to basic demographic characteristics such as sex and age, a broader range of clinical and contextual variables was collected to enrich patient profiling for the meta-analysis. These include physiological and lifestyle-related factors such as pregnancy status and breastfeeding status. Furthermore, ADHD-specific clinical features were recorded, including levels of inattentiveness, hyperactivity, and overall severity of symptoms.

- **Patient and clinician preferences**

On the other hand, a selection of side effects and clinical measures commonly associated with ADHD treatments was identified and included in the decision-making system. These features were rated by patients and doctors to incorporate their preferences into the multi-criteria evaluation process. The list of considered side effects and clinical measures includes Clinical Global Impression (CGI), drug use or misuse, seizures, accidents, insomnia, dizziness, vomiting, dry mouth, and vasovagal syncope.

Tables 6 and 7 show the preferences scores from patients and clinicians, respectively. All preferences were rated as highly important (≥ 7) by at least one patient, while clinicians rated some side effects, like dry mouth and tics, less critically. On the other hand, preferences like ADHD symptoms, CGI, seizures, and QoL consistently received high scores, reflecting strong agreement on their clinical relevance.

Table 6. Summary table of the preferences rating by simulated patients

	Adhd_sympmt_clin	CGI	Drug_use_prevention	Drug_use	Accidents	Academic_achievemnt	Appetite_decreased	Insomnia	Somnolence	Dizziness	Vomiting	Dry_mouth	Tic	Seizures	Vasovagal_syncope	QoL	N_dropd_ae	N_drop
count	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26
Mean	8.6	7.1	5.8	6	6.2	7.3	4.8	6.3	6.2	6.6	6.8	5	6.5	7.8	7	7.8	6.9	7.2
Std	0.8	2.1	3.1	2.9	2.3	2.3	2.7	2.4	2.3	2.4	2.1	2.1	2.1	1.7	2.1	1.5	2.2	2.2
Min	6	2	1	1	1	1	1	1	1	2	2	1	2	3	3	5	1	2
Max	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

Table 7. Summary table of the preferences rating by simulated clinicians

	Adhd_sympmt_clin	CGI	Drug_use_prevention	Drug_use	Accidents	Academic_achievemnt	Appetite_decreased	Insomnia	Somnolence	Dizziness	Vomiting	Dry_mouth	Tic	Seizures	Vasovagal_syncope	QoL	N_dropd_ae	N_drop
count	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28
Mean	7.9	6.6	3	3.3	4.2	6.1	4.2	6.4	4.9	5.2	4.6	2.1	4.3	6.1	5.8	7.3	5.9	6.7
Std	1	1.1	2.3	2.3	1.5	2.5	2.3	1.5	1.8	1.5	1.7	1	2	1.4	1.6	1	1.1	1.3
Min	6	3	1	1	1	1	1	2	1	2	1	1	1	3	1	5	2	5
Max	9	9	9	8	7	9	8	8	8	9	9	5	6	9	9	9	9	9

- **Interventions**

The final component incorporated into this decision-support system is the set of interventions identified through a comprehensive meta-analysis of existing clinical studies. For instance, atomoxetine, viloxazine, and guanfacine, among others.

5.1.2 EVALUATION METRICS OF THE PRELIMINARY EXPERIMENTS

In these first experiments, the evaluation consisted of the following metric:

1. Number of ties within the top N rankings.

In clinical practice, decision-making tools must minimize ambiguity. Ties within the top-ranking positions can hinder physicians from selecting the most suitable intervention. Since clinicians typically focus on the highest-ranking recommendations, we assessed the number of ties occurring within the top 2 to 5 positions. The number of ties at each level (e.g., top 3) was computed by subtracting the number of ties from the previous level (e.g., top 2) to provide a clearer breakdown.

2. Ranking of uniqueness and recommendation distribution

We calculated how many rankings were “clear or unequivocal”, meaning all recommended interventions had unique scores with no ties. For each MCDA method, we also computed statistical measures, including the mean, standard deviation, median, minimum, and maximum number of recommendations, along with the percentage of patients who received a perfect ranking. These results were compiled into a table for straightforward interpretation and comparison.

3. Percentage of patients receiving good ranking within the top N

This analysis checked whether the top N interventions appeared in exactly the first N positions without any tie or displacement. A binary score (1 or 0) was assigned per patient based on this condition, and the percentage was calculated across all patients for each top N subset (from 2 to 5).

5.1.3 RESULTS OF THE PRELIMINARY EXPERIMENTS

In these preliminary experiments, the evaluated approaches included the original APPRAISE-RS-E methodology and the enhanced recommender system, each time integrating one of the MCDA methods: WSM, COPRAS, TOPSIS or MOORA.

The results clearly demonstrate that APPRAISE-RS-E produces the highest number of total ties (511), including 72 ties at the Top 2 level alone. In contrast, all the enhanced MCDA-based systems exhibit a marked reduction in the total number of ties, with WSM showing 461 and the other three each producing only 460. Notably, COPRAS, MOORA and TOPSIS achieved zero ties at the Top 2 level, thereby eliminating ambiguity at the highest recommendation level. Although WSM reduced ties overall, it still retained a single tie at the Top 2 level and showed a higher number of ties at Top3 and Top 4 positions compared to the other methods.

Table 8. Tie distribution across different ranking positions in the preliminary experiments

MCDA	Top 2-ties	Top3- ties	Top4-ties	Top5- ties	Remaining ties	Total ties
APPRAISE-RS-E	72	8	35	28	368	511
WSM	1	33	63	6	358	461

TOPSIS	0	34	34	33	359	460
MOORA	0	32	62	1	365	460
COPRAS	0	32	62	1	365	460

Regarding metric about the ranking of uniqueness and recommendation distribution, the results show that the new recommender system incorporating any of the MCDA methods achieved higher mean (≈ 17.07) and median (20.5) values compared to APPRAISE-RS-E (mean: 15.11, median: 16). On the other hand, these new recommender systems also outperformed APPRAISE-RS-E in discriminating treatments within the top-ranked positions. While APPRAISE-RS-E achieved 82,1% of receiving “clear or unequivocal” ranking in the Top 2, COPRAS and MOORA reached 92,9%, maintaining this performance through the Top 3, Top 4 and Top 5 with significantly higher percentages (e.g. in top 5 vs. 32,1% for APPRAISE-RS-E).

5.2 EXPERIMENTS WITH THE FINAL STRUCTURE OF THE RECOMMENDER SYSTEM ON THE FINAL VALIDATED DATASET

To better achieve the main objective of this work, delivering more personalized and discriminative treatment recommendations, additional enhancements were implemented beyond the previously mentioned modifications. These included separating the rankings for efficacy and safety preferences, increasing the granularity of the benefit/risk-quality scores and introducing an additional decision-making criterion: convenience into the recommender system.

As previously commented, a final dataset was constructed, incorporating 65 real patients in addition to the simulated patients. This dataset was used in the following experiments to generate more realistic and clinically meaningful outputs from the recommender system. Furthermore, interventions labeled as “No evidence” in the benefit/risk-quality judgments were penalized with a score of -1. Interventions without supporting evidence from existing studies were excluded entirely, which led to the removal of four patients from the analysis, resulting in a final dataset of 89 patients.

In the following experiments, we also introduced two new approaches: modified versions of TOPSIS and MOORA, named TOPSIS_2 and MOORA_2.

A known limitation of distance-based MCDA methods is their susceptibility to rank reversal. This phenomenon occurs when minor modifications to the input data, such as slight changes in benefit-risk assessments or preferences, lead to significant shifts in the final ranking of alternatives. To address this issue within our framework, we implemented a corrective strategy involving the manual specification of the PIS and NIS, representing the best and worst possible outcome across all criteria, prior to applying the distance-based MCDA techniques.

Specifically, we assigned a benefit/risk-quality score of +5 to each criterion in the PIS and -5 to the corresponding criteria in the NIS for efficacy preferences. However, for the safety preferences, the PIS was set to 0 and NIS to -5. This distinction reflects the conceptual difference in the objective: while efficacy aims to maximize positive clinical outcomes, the goal of safety variables is not simply to reduce adverse events below placebo levels, but rather to achieve a safety profile comparable to that of a placebo.

Since a placebo is pharmacologically inert and, theoretically, should not cause any adverse effects. The ideal scenario in terms of safety is when the frequency of adverse events is equal to that observed with a placebo. This justifies the assignment of 0 as the ideal value for safety variables.

These values represent the theoretical extremes of the benefit-risk spectrum for any given preference. By anchoring the analysis with these fixed reference points, the model becomes more resistant to small input variations, thereby enhancing the stability and reliability of the final rankings.

5.2.1 REAL AND SIMULATED PATIENT DATASET

Following the inclusion of the new patients, we observed a continued gender imbalance, with 65% male and 35% female. Similarly, the age distribution remained skewed toward the adolescent population, followed by the childhood. Consistent with the first dataset, we also gathered additional demographic variables such as employment status, pregnancy, and ADHD-specific clinical features like hyperactivity, severity, and comorbidities, enriching the dataset for more nuanced analysis.

In the second experiment, the number of interventions extracted from existing clinical studies increased from 36 to 65. The newly added interventions are primarily psychotherapeutic approaches, such as mindfulness, parent- and patient-focused behavioral therapy, and patient-focused relaxation therapy, among others.

ADHD patients often experience difficulties in interpersonal relationships, both at home and in the workplace, which can lead to criticism and social rejection. Skilled therapists can play a key role in helping patients process past experiences and improve their social and emotional functioning. Moreover, psychotherapies typically have fewer side effects than pharmacological treatments [54]. However, due to a lack of extensive clinical investigation compared to medications, these interventions may offer less robust evidence, potentially limiting their influence in the decision-making system.

Additionally, while the set of preferences remained largely unchanged, they were categorized into two groups: efficacy-related and safety-related preferences, as illustrated in Figure 2. This separation was introduced to facilitate a more structured computation of the final rankings, particularly when incorporating the convenience scores. Furthermore, two additional independent

datasets, one related to cost and the other to convenience, were integrated into the system. These components are described in greater detail in the development section.

Moreover, we introduced a convenience ranking, based solely on fixed drug characteristics and not dependent on individual patient profiles. The convenience score was calculated using four sub-criteria: (1) frequency of follow-up evaluations, categorized as every 3 months, every 6 months, more than 6 months or not applicable (N/A); (2) the number of pills per day, ranging from 1 to 3; (3) availability of non-solid formulations (yes or no); and (4) food-related considerations (yes or no). Each sub-criterion contributes equally, with a maximum of 0.25 points, leading to a maximum total convenience score of 1.0 for each intervention.

However, psychotherapeutic interventions are handled differently. Due to the lack of standardized convenience data, we assign them a fixed score of 0.125, the minimum score observed for pharmacological treatments. This decision is based on the fact convenience cannot be accurately measured for psychotherapies, given their wide variability, such as telematic formats, group vs. individual settings, number of sessions, asynchronous modalities, etc. Nonetheless, we choose not to assign a zero score to psychotherapies to avoid introducing a systematic bias against non-pharmacological options, which could unfairly penalize them in the final ranking.

5.2.2 EVALUATION METRICS

In these second round of experiments, the evaluation metrics were:

1. **Number of ties within the top N rankings.**
2. **Percentage of patients receiving “clear or unequivocal” ranking within the top N**
3. **Spearman rank correlation coefficient**

This metric measures the strength and direction of the monotonic relationship between pairs of rankings to evaluate the level of agreement between two rankings. It was used to analyze the robustness of the distance-based MCDA methods TOPSIS and MOORA, due to their known susceptibility to the rank reversal phenomenon. In this project, a high Spearman coefficient (for instance, >0.8) suggests they generate very similar rankings and a low Spearman coefficient (for example, <0.4) would imply the rankings diverge significantly.

4. **Clinical analysis**

This variable evaluates how frequently specific interventions appear among the top recommendations. It focuses on methylphenidate, lisdexamfetamine, atomoxetine, guanfacine, mixed amphetamine salts, viloxazine, paroxetine, patient-focused cognitive behavioral therapy, patient-focused neurocognitive training, and mindfulness. For each intervention, its frequency was calculated as the proportion of patients for whom it

appeared in the top recommendations. This analysis serves to determine whether the interventions recommended are in line with clinical practice. In this line, it is expected psychostimulant drugs (methylphenidate or lisdexamfetamine) to be ranked at the top more frequently than viloxazine or paroxetine.

5.2.3 RESULTS OF THIS SECOND ROUND OF EXPERIMENTS

Once the final dataset was generated and clinically validated, we tested our solution to assess the impact of both the initial and newly implemented improvements to the APPRAISE-RS-E methodology.

As shown in Table 9, WSM showed the highest number of ties in the top 2 positions, with 20 instances, making it less discriminative in the most critical ranking zone. MOORA, in both of its evaluated versions, followed with 5 ties within the top 2 recommendations. In contrast, the remaining methods demonstrated similar and favorable performance by placing the majority of their ties outside the top 5. This is particularly advantageous, as clinicians generally focus their attention on the top 5 recommendations when making treatment decisions.

Table 9. Number of ties within the top-ranked positions of recommender system on the final dataset when each MCDA method is integrated individually

MCDA	Top 2-ties	Top3- ties	Top4-ties	Top5- ties	Remaining ties	Total ties
WSM	20	0	0	0	139	159
TOPSIS	1	0	0	0	141	142
MOORA	5	1	5	19	112	142
COPRAS	1	0	0	0	141	142
TOPSIS_2	1	0	0	0	141	142
MOORA_2	5	2	3	10	122	142

Regarding the recommendations provided, there are no substantial differences between the aggregation techniques, as illustrated in Table A3. The mean number of recommendations ranges from 14 to 15, accompanied by a relatively high standard deviation of approximately 9 to 10. On the other hand, the differences between the aggregation methods are relatively minor, generally below 10%, as shown in Table 10. TOPSIS and COPRAS exhibit slightly better performance, with a higher percentage of patients showing uniquely ranked interventions in the top 2 and 3 positions. However, more noticeable variations appear in the top 4 and 5 ranges. Similarly, WSM and MOORA display lower percentages but follow the same general pattern.

Table 10. Percentage of patients receiving “clear or unequivocal” recommendations within top-x rankings of recommender system on the final dataset when each MCDA method is integrated individually

MCDA	Top 2 (%)	Top 3 (%)	Top 4 (%)	Top 5 (%)
WSM	88.764	86.517	80.899	70.786
TOPSIS	92.135	88.764	83.146	73.034
COPRAS	92.135	88.764	83.146	73.034
MOORA	88.764	84.270	78.652	66.292
TOPSIS_2	92.135	88.764	83.146	73.034
MOORA_2	88.764	84.270	77.528	67.416

As commented in the design section, distance-based MCDA methods are prone to rank reversal, where small changes can significantly affect the output ranking. To assess robustness, two manually defined ideal solutions (PIS and NIS) were introduced into the dataset. We then re-executed the full system pipeline and calculated the Spearman rank correlation coefficients between the original and modified rankings. TOPSIS achieved a correlation of 0.91, while MOORA scored 0.83. These results suggest that TOPSIS is more robust than MOORA when ideal solutions are altered.

In previous evaluations, we did not exclude interventions with an unfavorable benefit-risk balance, focusing instead on whether the rankings included the full set of available interventions. From this moment on we adopt a more clinically oriented perspective by filtering for favorable clinical recommendations only. This allows us to better assess the clinical utility of the system. We focus on TOPSIS and COPRAS due to their strong performance in earlier experiments and the interpretability of their results.

In the case of COPRAS, this method operates using similar arithmetic principles to compute the overall performance score for each alternative. The score is calculated through weighted summation, where each alternative’s normalized performance values are multiplied by the corresponding criterion weights and then aggregated. This final score determines the alternative’s position in the ranking, Equation 9.

After calculating separate rankings based on patient and clinician preferences, a weighted summation method is applied to merge them, giving a weight of 0.30 to the patient and 0.70 to the clinician, following the APPRAISE-RS-E framework.

To validate and adjust the merged results, an integration of in favor and against rankings is performed. In this step, the favor score is subtracted by the against score, Equation 14.

$$\text{Final Score} = \text{Favor Score} - \text{Against Score}$$

Equation 14. Calculation of the final score by merging favor and against rankings

This ensures that if an intervention has a predominantly negative benefit-risk assessment, its final score reflects that negativity, even if it performed well in preference-based rankings. This approach allows the system to preserve alignment with clinical judgment and evidence and helps prevent recommending interventions that may not be favorable overall.

On the other hand, the TOPSIS method calculates the final score of each intervention based on its Euclidean distance from two reference points: the PIS and the NIS, as shown in Equations 10-12.

The idea behind TOPSIS is intuitive, the best alternative should be as close as possible to the ideal solution and as far as possible from the least desirable one.

For example:

- An ideal intervention would have a distance to the PIS (S_{i+}) of approximately 0.001 and a distance to the NIS (S_{i-}) of approximately 0.999. This results in a final score close to 0.999, indicating a highly favorable alternative.
- Conversely, the least favorable intervention would yield $S_{i+} \approx 0.999$ and $S_{i-} \approx 0.001$, producing a final score around 0.001, reflecting poor performance across key criteria.

This score is then incorporated into the integration of in favor and against rankings, where the favor score is subtracted by the against score. If the result is positive, it implies that the intervention leans towards a favorable benefit-risk profile; if negative, it suggests a preference toward the placebo or an unfavorable profile.

This is only possible when the ideal solution is defined manually because if it is derived directly from the dataset, it corresponds to the maximum score observed for each preference, not necessarily the theoretical maximum. As a result, the computed ideal solution may not always reflect the most optimal or extreme possible values, but rather the best available within the given data.

After applying the filtering criteria, 58 patients remained for the COPRAS method, while TOPSIS_2 retained 57 patients. Based on the updated rankings, COPRAS provided an average of 2.29 recommendations per patient (SD = 1.22), while TOPSIS_2 offered an average of 3.46 recommendations (SD = 2.37). The median and maximum number of recommendations were 2 and 6 for COPRAS, and 3 and 11 for TOPSIS_2, respectively.

Then, a clinical analysis was performed to identify the most commonly recommended options with respect the selected interventions analyzed across diverse patient profiles.

As shown in Table 11, methylphenidate was consistently the most frequently recommended intervention, appearing as the top-ranked option in 63.8% (COPRAS) and 57.9% (TOPSIS_2) of cases. It maintained a strong presence across the top two and top three positions in both methods. Atomoxetine followed, with top three appearances in 77.6% (COPRAS) and 75.4% (TOPSIS_2) of cases, while lisdexamfetamine showed moderate frequency, reaching top three positions in over 40% of cases for both methods.

In contrast, other pharmacological options like guanfacine and paroxetine, as well as non-pharmacological interventions such as cognitive behavioral therapy, neurocognitive training, and mindfulness, did not appear among the top three recommendations in any case. Mixed amphetamine salts and viloxazine showed limited relevance, appearing in top three rankings for fewer than 10% of patients. These results were interpreted by the clinicians in the project team as evidence supporting the clinical validity of the system performance.

Table 11. Frequency of appearance of pharmacological and psychotherapeutic interventions in top 3 ranking positions across the recommender system on the final dataset when TOPSIS (T2) and COPRAS (C) methods are integrated individually. In the case of TOPSIS, PIS and NIS are manually introduced.

Intervention	Top 1-C	Top 1-T2	Top 2-C	Top 2-T2	Top 3-C	Top 3-T2
Methylphenidate	63.79	57.89	75.86	66.67	75.86	70.17
Lisdexamfetamine	6.89	8.77	36.21	29.82	41.38	40.35
Atomoxetine	29.31	28.07	50.0	49.12	77.59	75.43
Guanfacine	0	0	0	0	0	0
Mixed amphetamine salts	0	0	6.90	0	10.34	3.51
Viloxazine	0	1.75	0	1.75	0	3.51
Paroxetine	0	0	0	0	0	0
Patient focused cognitive behavioral therapy	0	0	0	0	0	0
Patient focused neurocognitive training	0	0	0	0	0	0
Mindfulness	0	0	0	0	0	0

5.3 FINAL ANALYSIS

In the final phase of this project, our objective was to directly compare the newly recommender system, integrating the best-performing MCDA method, with the baseline methodology we aimed to improve: APPRAISE-RS-E.

To achieve this, we have selected the MCDA methods which demonstrated the highest performance in terms of discrimination capability: the discrimination provided: TOPSIS_2 and COPRAS, and also the APPRAISE-RS-E which is the background method in which several enhancements were performed. The evaluation was conducted using a dataset composed exclusively of real patients' cases. Two key evaluation metrics were applied:

1. Number of ties within the top N treatments rankings, to assess ranking discrimination.
2. Percentage of patients receiving a good ranking within the top N positions, to evaluate recommendations quality.

As shown in Table 12, APPRAISE-RS-E shows a very high number of ties, particularly in the Top 2 (466 ties) and Top 5 (368 ties) recommendations, with nearly 2000 total ties across the rankings. In contrast, both COPRAS and TOPSIS_2 show only one tie in the Top 2 and a total of just 20 ties.

Table 12. Number of ties within the top-ranking recommendations of the final recommender system when TOPSIS_2 and COPRAS are integrated individually and APPRAISE-RS-E

MCDA	Top2 ties	Top3 ties	Top4 ties	Top5 ties	Remaining ties	Total ties
APPRAISE-RS-E	466	78	133	368	904	1949
COPRAS	1	0	0	0	19	20
TOPSIS_2	1	0	0	0	19	20

Table 13 shows that APPRAISE-RS-E achieves good ranking for only 64% of patients in the Top 2 recommendations, and the percentage drops sharply to just 15,63% by the Top 5. As opposed, both COPRAS and TOPSIS_2 maintain a very high quality of recommendation, with over 95% of patients receiving at least one unique treatment in the Top 2, and nearly 70% even by the Top 5.

Table 13. Percentage of patients receiving "clear or unequivocal" recommendations ranking of the new recommender system when TOPSIS_2 and COPRAS are integrated individually and APPRAISE-RS-E

MCDA	Top 2 (%)	Top 3 (%)	Top 4 (%)	Top 5 (%)
APPRAISE-RS-E	64.062	39.062	23.437	15.625
COPRAS	95.082	88.525	83.606	68.852
TOPSIS_2	95.082	88.525	83.606	68.852

6. DISCUSSION

The main objective of this project, outlined in detail in the Objectives section, was to provide more personalized and discriminative treatment recommendations for ADHD patients, improving upon the original APPRAISE-RS-E methodology.

In the preliminary evaluations, the APPRAISE-RS-E methodology produced a high number of tied recommendations compared to the enhanced recommender system. This was largely due to its design constraints, only preferences rated 7 or higher were considered and weightings of the benefit/risk-quality judgements were simplified to two levels. These limitations reduced variability across patients and interventions, ultimately diminishing personalization.

In contrast, the enhanced system integrated all patient and clinician preferences and increased the granularity of benefit/risk-quality scores and weights. Additionally, the adoption of alternative MCDA methods significantly improved ranking differentiation. Notably, all simulated patients received a unique top-ranked intervention, in contrast to the 72 top ties observed with APPRAISE-RS-E. This improvement is clinically significant, as it reduces ambiguity in decision-making and enables clinicians to more confidently identify the most appropriate treatment for each patient.

In the second round of experiments with real and simulated patients, a more diverse and complex profile led to an increase in standard deviation and the total number of ties, especially in the middle of rankings. While TOPSIS and COPRAS showed strong performance with only one tie in the top five for any patient, the total number of ties across the full ranking reached 141, indicating challenges in differentiation among mid-tier options. These middle ties do not significantly impact clinical decision-making. Clinicians primarily rely on the top five ranked interventions, rendering ties beyond this threshold less relevant for practical use.

Although MOORA also demonstrated strong performance, it posed challenges in interpreting the final scores. Specifically, we were unable to establish a reliable threshold to distinguish between favorable and unfavorable benefit-risk balances. This ambiguity would make it difficult to effectively communicate the method's rationale to clinicians, as its score range lacked clear clinical meaning or comparability.

Focusing on the most recommended interventions from the best-performing methods, COPRAS and the updated TOPSIS (TOPSIS_2), methylphenidate, lisdexamfetamine, and atomoxetine consistently ranked within the top 1 to 3 positions. This aligns with their worldwide use (including Spain) for ADHD treatment, supported by strong clinical evidence [55]. As expected, paroxetine, typically prescribed for comorbid mood or anxiety disorders rather than ADHD itself, did not appear among the top recommendations [56].

Mixed amphetamine salts and viloxazine, although approved in the United States, have a less widespread research use with no approval for clinical use in Spain, which may explain their lower ranking [57, 58]. On the other hand, psychotherapeutic interventions, such as cognitive behavioral therapy and neurocognitive training, have gained research interest in recent years due to their potential to manage ADHD symptoms and improve quality of life. However, their limited clinical evidence in ADHD-specific contexts likely contributed to their absence from the top-ranked recommendations. [59]

When comparing the enhanced recommender system with the original APPRAISE-RS-E on real patient data, a clear improvement in discrimination was observed. The original method yielded 466 top two ties across 23 patients (36% of the real sample), with an average of 20 tied interventions out of 65 per patient. This underscores its limited utility in more complex cases and highlights the value of the proposed enhancements.

Beyond the technical and clinical contribution of this project, it is important to consider its broader ethical, sustainability and diversity implications, which are essential for responsible and equitable implementation in real-world healthcare settings.

From an ethical perspective, the recommender system promotes patient-centered care by empowering individuals in informed decision-making through the integration of their personal preferences into treatment selection. While the system aims to accommodate a wide range of ADHD cases, certain populations, such as patients over the age of 60, remain underrepresented. Future work should focus on enhancing algorithmic fairness by ensuring broader population coverage and by developing more robust methods to identify and mitigate potential biases from the dataset related to age, gender, or socioeconomic status.

In terms of sustainability, the system has the potential to improve clinical efficiency by supporting clinicians in the time-intensive process of treatment planning, thereby freeing up resources for managing more complex cases. Furthermore, the integration of a cost-related variable can empower patients to consider both clinical effectiveness and economic feasibility when selecting treatments, reducing financial burden and promoting more sustainable long-term care decisions.

Regarding diversity and inclusion, the dataset encompasses a wide range of ages, sexes, and clinical profiles, which enhances the system's applicability across diverse patient populations. Additionally, the inclusion of both pharmacological and non-pharmacological interventions promotes a more holistic and inclusive approach to ADHD care. Although cost was not directly incorporated into the decision-making algorithm, the system provides information on the monthly cost of each recommended drug, helping patients and caregivers make financially informed choices.

7. CONCLUSIONS

The primary goal of this project was to deliver more personalized and discriminative treatment recommendations. It was successfully achieved through several key enhancements: the integration of more granular benefit/risk-quality scores, an expanded range of their weighting in the decision-making process, incorporation of all patient and clinician preferences, evaluation of alternative MCDA techniques and the inclusion of the convenience as an additional decision-making criterion. These enhancements were particularly relevant in the context of ADHD, a condition characterized by diverse and complex patient profiles, where individualized treatment is essential.

Regarding personalization, the incorporation of all patient and clinician preferences along with the inclusion of the convenience variable helped align treatment recommendations more closely with real-world patient needs. Factors like the number of pills or formulation type were considered, thereby increasing patient participation in treatment selection and potentially reducing treatment dropout. Although the cost variable could not be formally incorporated due to insufficient data available on non-pharmacological interventions (e.g. psychotherapies), it is planned to be included as a descriptive field in the visual ranking to support informed, cost-aware decision making.

To improve the system's discriminative power, the benefit/risk-quality scoring system was made more granular, and its weightings were refined. While several alternative MCDA methods were reviewed, only four were selected due to their simplicity and compatibility with the input data. These selected techniques were tested in place of the traditional WSM. These alternatives, especially TOPSIS and COPRAS, demonstrated superior performance by consistently producing a unique top-ranked intervention for all simulated and real patients, leading to a substantial reduction in ties among top-ranked interventions. This outcome highlights a known limitation of WSM: its vulnerability to distortion by heavily weighted criteria, which can overshadow other relevant factors and result in non-discriminative outputs.

Between the two TOPSIS versions, one with and one without the inclusion of predefined PIS and NIS, we selected the latter (TOPSIS_2) as the candidate method alongside COPRAS. This decision was based on the fact that TOPSIS_2 calculates the Euclidean distance using absolute maximum values rather than relative ones, a feature that proved crucial in accurately filtering favorable clinical recommendations.

In summary, the proposed recommender system improves both personalization and discrimination of treatment recommendations, addressing key limitations of existing approaches such as the APPRAISE-RS-E methodology and the Cochrane/GRADE heuristic. It provides clinicians with clearer, more individualized guidance, which can improve clinical decision-making and foster greater trust in the use of AI-supported tools in healthcare.

8. LIMITATIONS AND FURTHER DIRECTIONS

While the results of this project demonstrate clear improvement in treatment personalization and discrimination, several limitations should be acknowledged, and key directions for future work have been identified.

1. Generalizability and patient diversity

The current system was tested on a dataset compromising simulated and real patients, but the sample may not fully capture the heterogeneity present in the broader ADHD population. In particular, certain subgroups, such as elderly patients, individuals with multiple comorbidities, or patients from underrepresented socioeconomic backgrounds, were not adequately represented. Future work should focus on validating the system with a more diverse and representative population to ensure broader applicability and fairness.

2. Fixed assumptions in weighting preferences

In this version, the system assumes a fixed relative importance of 70% for clinician preferences and 30% for patient preferences. Similarly, a fixed distribution of 45%-45%-10% was applied to efficacy, safety and convenience, respectively. These predefined weights, while reasonable starting points, may not reflect all clinical scenarios or individual user values. Future modifications can consist of adjusting importance levels in real-time or offering predefined templates based on clinical context.

3. Cost integration

While drug cost is an important component of real-world decision-making, integrating it effectively into the MCDA framework proved challenging, especially due to limited cost data for non-pharmacological interventions. Future work should aim to improve the completeness and accuracy of cost data and explore ways to meaningfully incorporate it as either a formal criterion or an adjustable constraint, without compromising recommendations quality.

4. Clinical validation

Although the system was tested with clinically informed datasets, formal validation in real clinical settings remains pending. A prospective validation study involving real clinicians and patients would provide critical evidence regarding the tool's clinical utility, usability and impact on decision-making and patient outcomes.

5. System integration and implementation

A key next step is the integration of the recommender system into clinical information systems and eventually into a full decision support system. This would involve addressing issues related to interoperability, data privacy, user interface design and workflow integration. In particular, the proposed solution could be embedded within the existing Evimatic platform, an information system already in use for evaluating the benefit-risk profile of ADHD treatments [25]. Evimatic is freely accessible and widely adopted, especially within the mental health network, where it is available through the electronic health record. Incorporating the recommender system into this established infrastructure would facilitate its clinical adoption and promote its practical application in routine care.

6. Scalability to other health conditions

While this project focused on ADHD, the underlying MCDA framework and system architecture are adaptable to other chronic and complex health conditions. Future work could involve scaling the system to address treatment decisions in areas such as depression, chronic pain, or cardiovascular disease, where individualized cases are equally important.

9. REFERENCES

- [1]: I. Ratnani, S. Fatima, M. M. Abid, Z. Surani, and S. Surani, "Evidence-based medicine: history, review, criticisms, and pitfalls," *Cureus*, vol. 15, no. 2, p. e35266, 2023. [Online]. Available: <https://doi.org/10.7759/cureus.35266>
- [2]: S. Tenny and M. A. Varacallo, "Evidence-Based Medicine," *StatPearls*, Sep. 2024, Accessed: Apr. 16, 2025. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK470182/>
- [3]: R. Busse, N. Klazinga, D. Panteli, and W. Quentin, "Improving healthcare quality in Europe," *European Observatory on Health Systems and Policies*, 2019, Accessed: Jun. 05, 2025. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK549276/>
- [4]: E. Guerra-Farfan, Y. Garcia-Sanchez, M. Jornet-Gibert, J. H. Nuñez, M. Balaguer-Castro, and K. Madden, "Clinical practice guidelines: The good, the bad, and the ugly," *Injury*, vol. 54, pp. S26–S29, May 2023, doi: 10.1016/J.INJURY.2022.01.047
- [5]: "Overview of clinical practice guidelines - UpToDate." Accessed: Apr. 16, 2025. [Online]. Available: <https://www.uptodate.com/contents/overview-of-clinical-practice-guidelines#H1479879854>
- [6]: Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. *Clinical practice guidelines we can trust*. Washington, DC: National Academies Press; (2011). Available from: <https://doi.org/10.17226/13058>
- [7]: Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Schünemann HJ, Edejer T, Varonen H, Vist GE, Williams JW Jr, Zaza S; GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ*. (2004) 328(7454):1490. doi: 10.1136/bmj.328.7454.1490
- [8]: Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, Treweek S, Mustafa RA, Vandvik PO, Meerpohl J, Guyatt GH, Schünemann HJ; GRADE Working Group. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. *BMJ*. (2016) 353:i2089. doi: 10.1136/bmj.i2089. PMID: 27365494
- [9]: A. Espinosa Brito, L. del Sol Padrón, A. Espinosa Brito, J. Garriga Valdés, and B. Viera Valdés, "Guías de práctica clínica. Ventajas y desventajas: Una propuesta de indicadores," *MediSur*, vol. 7, no. 5, pp. 44–47, 2009, Accessed: Apr. 16, 2025. [Online]. Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1727-897X2009000500009&lng=es&nrm=iso&tng=es
- [10]: S. H. Woolf, R. Grol, A. Hutchinson, M. Eccles, and J. Grimshaw, "Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines," *BMJ (Clinical research ed.)*, vol. 318, no. 7182, pp. 527–530, 1999. [Online]. Available: <https://doi.org/10.1136/bmj.318.7182.527>
- [11]: J. V. A. Franco, M. Arancibia, N. Meza, E. Madrid, and K. Kopitowski, "Clinical practice guidelines: Concepts, limitations and challenges," *Medwave*, vol. 20, no. 3, Apr. 2020, doi: 10.5867/MEDWAVE.2020.03.7887.
- [12]: "GRADE handbook." Accessed: Apr. 16, 2025. [Online]. Available: <https://gdt.gradepro.org/app/handbook/handbook.html#h.f06pwvbsyasx>
- [13]: Montori VM. Removing the blindfold: The centrality of care in caring for patients with multiple chronic conditions. *Health Serv Res*. (2021) 56 Suppl 1:969-972. doi: 10.1111/1475-6773.13865.
- [14]: Barry HC, Cosgrove L, Slawson DC. Where Clinical Practice Guidelines Go Wrong. *Am Fam Physician*. (2022) 105:350-352.
- [15]: P. G. Shekelle, E. Ortiz, S. Rhodes, S. C. Morton, M. P. Eccles, J. M. Grimshaw, and S. H. Woolf, "Validity of the Agency for Healthcare Research and Quality clinical practice guidelines: how quickly

- do guidelines become outdated?," JAMA, vol. 286, no. 12, pp. 1461–1467, 2001. [Online]. Available: <https://doi.org/10.1001/jama.286.12.1461>
- [16]: R. W. van Hamersvelt, M. Zreik, M. Voskuil, M. A. Viergeever, I. Išgum, and T. Leiner, "Deep learning analysis of left ventricular myocardium in CT angiographic intermediate-degree coronary stenosis improves the diagnostic accuracy for identification of functionally significant stenosis," *European Radiology*, vol. 29, no. 5, pp. 2350–2359, May 2019, doi: 10.1007/S00330-018-5822-3.
- [17]: P. Siciarz, S. Alfaifi, E. van Uytven, S. Rathod, R. Koul, and B. McCurdy, "Machine learning for dose-volume histogram based clinical decision-making support system in radiation therapy plans for brain tumors," *Clinical and Translational Radiation Oncology*, vol. 31, pp. 50–57, Nov. 2021, doi: 10.1016/J.CTRO.2021.09.001.
- [18]: "View of Integrating Artificial Intelligence into Clinical Practice Guidelines: Opportunities and Challenges." Accessed: Apr. 16, 2025. [Online]. Available: <https://www.journal.ypidathu.or.id/index.php/midwifery/article/view/1216/843>
- [19]: B. López et al., "APPRAISE-RS: Automated, updated, participatory, and personalized treatment recommender systems based on GRADE methodology," *Heliyon*, vol. 9, no. 2, Feb. 2023, doi: 10.1016/j.heliyon.2023.e13074.
- [20]: O. Raya, "Management of patient and physician preferences and explanations for participatory evaluation of treatment with an ethical seal." [Online]. Available: https://caleta.udg.edu/git/eXiT_Research_Group/TDAH_
- [21]: Evgenia Baykova, Òscar Raya, Cristina Lombardía, Begoña Gonzalvo, Inés Andreu, David Losada, Tania Falkenhain, Ruth Cunill, Domènec Serrano, David Rigau, David Ramírez, Beatriz López, Xavier Castells. Assessing TDAp: an AI-Based Clinical Decision Support System for ADHD Treatment Recommendations (under review)
- [22]: Oscar Raya; Xavier Castells; David Ramírez; Beatriz López. Patients and clinicians preferences together in the loop for ADHD treatment recommendations. Modeling Decisions for Artificial Intelligence 2023
- [23]: T. P. G. M. de Vries, R. H. Henning, H. V. Hogerzeil, and D. A. Fresle, *Guide to Good Prescribing: A Practical Manual*, Geneva: World Health Organization, WHO/DAP/94.11, 1994
- [24]: European Parliament and Council of the European Union, Regulation (EU) 2017/745 on medical devices, *Official Journal of the European Union*, L 117/1–L 117/175, Apr. 5, 2017.
- [25]: "evimatic | the first eHealth tool for ADHD." Accessed: Apr. 16, 2025. [Online]. Available: <https://evimatic.com/es/>
- [26]: Tanios N, Wagner M, Tony M, Baltussen R, van Til J, Rindress D, Kind P, Goetghebeur MM. Which Criteria Are Considered in Healthcare Decisions? Insights From an International Survey of Policy and Clinical Decision Makers. *Int J Technol Assess Health Care*. 2013; 29(4):456–65. <https://doi.org/10.1017/S0266462313000573>
- [27]: S. Chakraborty, R. D. Raut, T. M. Rofin, and S. Chakraborty, "A comprehensive and systematic review of multi-criteria decision-making methods and applications in healthcare," *Healthcare Analytics*, vol. 4, p. 100232, Dec. 2023, doi: 10.1016/J.HEALTH.2023.100232.
- [28]: M. A. Deif, A. A. A. Solyman, M. H. Alsharif, and P. Uthansakul, "Automated triage system for intensive care admissions during the COVID-19 pandemic using hybrid XGBoost-AHP approach," *Sensors (Basel, Switzerland)*, vol. 21, no. 19, p. 6379, 2021. [Online]. Available: <https://doi.org/10.3390/s21196379>
- [29]: R. Pramanik, S. Dey, S. Malakar, S. Mirjalili, and R. Sarkar, "TOPSIS aided ensemble of CNN models for screening COVID-19 in chest X-ray images," *Scientific Reports*, vol. 12, no. 1, p. 15409, 2022. [Online]. Available: <https://doi.org/10.1038/s41598-022-18463-7>

- [30]: T. Singh, C. Goswami, A. Patnaik, and L. Lendvai, "Optimal design of ceramic based hip implant composites using hybrid AHP-MOORA approach," *Materials* (Basel, Switzerland), vol. 15, no. 11, p. 3800, 2022. [Online]. Available: <https://doi.org/10.3390/ma15113800>
- [31]: H. Ansari pour, K. L. Haeussler, S. J. Ferguson, and M. Flohr, "Prioritizing biomaterials for spinal disc implants by a fuzzy AHP and TOPSIS decision making method," *Scientific Reports*, vol. 13, no. 1, p. 21531, 2023. [Online]. Available: <https://doi.org/10.1038/s41598-023-48735-9>
- [32]: C. Türkmen, M. Konca, and B. Yetim, "Prioritization of neurologic rehabilitation interventions by ELECTRE-III analysis in subacute stroke patients," *Acta Neurologica Belgica*, vol. 123, no. 1, pp. 181–189, 2023. [Online]. Available: <https://doi.org/10.1007/s13760-022-01982-5>
- [33]: B. Meniz and E. M. Özkan, "Vaccine selection for COVID-19 by AHP and novel VIKOR hybrid approach with interval type-2 fuzzy sets," *Engineering Applications of Artificial Intelligence*, vol. 119, p. 105812, Mar. 2023, doi: 10.1016/J.ENGAPPAI.2022.105812.
- [34]: Poehler D, Czerniecki J, Norvell D, Henderson A, Dolan J, Devine B. The Development and Pilot Study of a Multiple Criteria Decision Analysis (MCDA) to Compare Patient and Provider Priorities around Amputation-Level Outcomes. *MDM Policy & Practice*. 2022;7(2). doi:10.1177/23814683221143765
- [35]: I. Ozsahin, S. T. Abebe, and G. S. P. Mok, "A multi-criteria decision-making approach for schizophrenia treatment techniques," *Archives of Psychiatry and Psychotherapy*, vol. 22, no. 2, pp. 52–61, 2020. [Online]. Available: <https://doi.org/10.12740/APP/111624>
- [36]: I. Ozsahin, M. T. Mustapha, S. Albarwary, B. Sanlidag, D. U. Ozsahin, and T. A. Butler, "An investigation to choose the proper therapy technique in the management of autism spectrum disorder," *Journal of Comparative Effectiveness Research*, vol. 10, no. 5, pp. 423–437, 2021. [Online]. Available: <https://doi.org/10.2217/cer-2020-0162>
- [37]: P. Majumder, A. Pal, D. R. Dorai et al., "Accelerating depression intervention: identifying critical psychological factors using MCDM-MOORA technique for early therapy initiation," *Annals of General Psychiatry*, vol. 23, no. 35, 2024. [Online]. Available: <https://doi.org/10.1186/s12991-024-00518-w>
- [38]: "Advantages And Limitations Of The Weighted Sum Method - FasterCapital." Accessed: Apr. 18, 2025. [Online]. Available: <https://fastercapital.com/topics/advantages-and-limitations-of-the-weighted-sum-method.html>
- [39]: "► Weighted sum model: Simple explanation and example." Accessed: Apr. 18, 2025. [Online]. Available: <https://www.munich-business-school.de/en/l/business-studies-dictionary/weighted-sum-model>
- [40]: "skcriteria.madm.simple module — scikit-criteria 0.8 documentation." Accessed: Apr. 18, 2025. [Online]. Available: <https://scikit-criteria.quatrope.org/en/0.8/api/madm/simple.html#skcriteria.madm.simple.wpm>
- [41]: S. B. Patil, T. A. Patole, R. S. Jadhav, S. S. Suryawanshi, and S. J. Raykar, "Complex Proportional Assessment (COPRAS) based Multiple-Criteria Decision Making (MCDM) paradigm for hard turning process parameters," *Materials Today: Proceedings*, vol. 59, pp. 835–840, Jan. 2022, doi: 10.1016/J.MATPR.2022.01.142.
- [42]: A. Bilbao-Terol, M. Arenas-Parra, V. Cañal-Fernández, and J. Antomil-Ibias, "Using TOPSIS for assessing the sustainability of government bond funds," *Omega*, vol. 49, pp. 1–17, Dec. 2014, doi: 10.1016/J.OMEGA.2014.04.005.
- [43]: B. Yang, J. Zhao, and H. Zhao, "A robust method for avoiding rank reversal in the TOPSIS," *Computers & Industrial Engineering*, vol. 174, p. 108776, Dec. 2022, doi: 10.1016/J.CIE.2022.108776.
- [44]: P. Karande and S. Chakraborty, "Application of multi-objective optimization on the basis of ratio analysis (MOORA) method for materials selection," *Materials & Design*, vol. 37, pp. 317–324, May 2012, doi: 10.1016/J.MATDES.2012.01.013.

- [45]: F. D. Ahmed and M. A. Majid, "Towards agent-based petri net decision making modelling for cloud service composition: A literature survey," *Journal of Network and Computer Applications*, vol. 130, pp. 14–38, Mar. 2019, doi: 10.1016/j.jnca.2018.12.001.
- [46]: T. L. Saaty, "How to make a decision: The analytic hierarchy process," *European Journal of Operational Research*, vol. 48, no. 1, pp. 9–26, Sep. 1990, doi: 10.1016/0377-2217(90)90057-I.
- [47]: E. T. Superior, "Enginyeria, O. Joseph, S. Shofade, and R. Slowinski, "Universitat Rovira i Virgili Considering hierarchical structure of criteria in ELECTRE decision aiding methods," 2011.
- [48]: M. Fontana and J. Muñuzuri Sanz, "Equation Chapter 1 Section 1 Proyecto Fin de Carrera Ingeniería Industrial Métodos de decisión multicriterio AHP y PROMETHEE aplicados a la elección de un dispositivo móvil."
- [49]: "Qué es y cómo se calcula el ROI de la empresa | Blog Anfix." Accessed: May 04, 2025. [Online]. Available: <https://www.anfix.com/blog/que-es-roi>
- [50]: "Guía clínica de La evaluación de la calidad de la evidencia y la graduación de la fuerza de las recomendaciones: el sistema GRADE - Fisterra." Accessed: Apr. 23, 2025. [Online]. Available: <https://www.fisterra.com/guias-clinicas/la-evaluacion-calidad-evidencia-graduacion-fuerza-recomendaciones-sistema-grade/>
- [51]: J. Olivier, W. L. May, and M. L. Bell, "Relative effect sizes for measures of risk," *Communications in Statistics - Theory and Methods*, vol. 46, no. 14, pp. 6774–6781, Mar. 2017, doi: <https://doi.org/10.1080/03610926.2015.1134575>.
- [52]: J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, N.J: L. Erlbaum Associates, 1988.
- [53]: "Guía clínica de La evaluación de la calidad de la evidencia y la graduación de la fuerza de las recomendaciones: el sistema GRADE - Fisterra." Accessed: Apr. 23, 2025. [Online]. Available: <https://www.fisterra.com/guias-clinicas/la-evaluacion-calidad-evidencia-graduacion-fuerza-recomendaciones-sistema-grade/>
- [54]: De Crescenzo, Franco, et al. "Pharmacological and Non-Pharmacological Treatment of Adults with ADHD: A Meta-Review." *Evidence Based Mental Health*, vol. 20, no. 1, 19 Dec. 2016, pp. 4–11, ebmh.bmj.com/content/20/1/4, <https://doi.org/10.1136/eb-2016-102415>.
- [55]: Grupo de trabajo de la Guía de Práctica Clínica sobre las Intervenciones Terapéuticas en el Trastorno por Déficit de Atención con Hiperactividad (TDAH). Guía de Práctica Clínica sobre las Intervenciones Terapéuticas en el Trastorno por Déficit de Atención con Hiperactividad (TDAH). Ministerio de Sanidad, Servicios Sociales e Igualdad. Instituto Aragonés de Ciencias de la Salud (IACS); 2017 Guías de Práctica Clínica en el SNS.
- [56]: "CIMA:: FICHA TECNICA PAROXETINA AUROBINDO PHARMA 20 MG COMPRIMIDOS RECUBIERTOS CON PELICULA EFG." Accessed: Jun. 03, 2025. [Online]. Available: https://cima.aemps.es/cima/dochtml/ft/66218/FT_66218.html
- [57]: U.S. Food and Drug Administration, "ADDERALL XR® (mixed salts of a single-entity amphetamine product) dextroamphetamine sulfate, dextroamphetamine saccharate, amphetamine aspartate monohydrate, amphetamine sulfate capsules, CII," Shire US Inc., 2013. [Online]. Available: https://www.accessdata.fda.gov/drugsatfda_docs/label/2013/021303s026lbl.pdf. [Accessed: 3 Jun. 2025].
- [58]: U.S. Food and Drug Administration, "QELBREE® (viloxazine extended-release capsules), for oral use," Supernus Pharmaceuticals, Inc., 2022. [Online]. Available: https://www.accessdata.fda.gov/drugsatfda_docs/label/2022/211964s003lbl.pdf. [Accessed: 3 Jun. 2025].

- [59]: Young, Susan, and J. Myanthi Amarasinghe. "Practitioner Review: Non-Pharmacological Treatments for ADHD: A Lifespan Approach." *Journal of Child Psychology and Psychiatry*, vol. 51, no. 2, Feb. 2010, pp. 116–133, onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-7610.2009.02191.x

10. ANNEXES

10.1 SUPPLEMENTARY TABLES

Table A1 Percentage of patients receiving “clear or unequivocal” recommendations within top-x rankings when a MCDA method is integrated individually in the first experiments

MCDA	Top 2 (%)	Top 3 (%)	Top 4 (%)	Top 5 (%)
APPRAISE-RS-E	82.143	64.286	39.286	32.143
WSM	92.857	89.286	82.143	64.286
TOPSIS	92.857	92.857	82.143	67.857
COPRAS	92.857	92.857	89.286	75.0
MOORA	92.857	92.857	89.286	75.0

Table A2. Statistical metrics of the number of recommendations per patient provided by recommender system when MCDA methods are integrated individually in the first experiments

MCDA	Mean	SD	Median	Minimum	Maximum
APPRAISE-RS-E	15.111	7,44	16	1	25
WSM	17.035	8.011	20.5	1	25
TOPSIS	17.071	8.044	20.5	1	25
COPRAS	17.071	8.044	20.5	1	25
MOORA	17.071	8.044	20.5	1	25

Table A3. Statistical metrics of the number of recommendations per patient provided by recommender system when MCDA methods are integrated individually in the second experiments

MCDA	Mean	SD	Median	Minimum	Maximum
WSM	14.89	10.151	13	1	36
TOPSIS	14.281	9.932	13	1	36
COPRAS	14.281	9.932	13	1	36
MOORA	14.281	9.932	13	1	36
TOPSIS_2	14.281	9.932	13	1	36
MOORA_2	14.281	9.932	13	1	36