

Prescriptive Analytics-based Methodologies for Healthcare Data: A Systematic Literature Review

Gustavo Emilio Mendoza-Olguín^{1,*}, María Josefa Somodevilla-García¹,
Concepción Pérez-de-Celis¹, Yanin Chavarri-Guerra²

¹ Benemérita Universidad Autónoma de Puebla, Puebla,
Mexico

² Instituto Nacional de Ciencias Médicas y Nutrición “Salvador Zubirán”, CDMX,
Mexico

{gustavo.mendozaol, maria.somodevilla, concepcion.perezdecelis}@correo.buap.mx,
yanin.chavarrig@incmnsz.mx

Abstract. This paper reviews the use of prescriptive approaches to develop personalized treatment plans in healthcare. A systematic literature review was conducted to identify existing methodologies developed for suggesting treatments in medicine using Systematic Mapping Process, PRISMA methodology, and recommended Critical Appraisal Tools. A final set of 30 documents was analyzed. The findings indicate that despite efforts to establish prescriptive methodologies, there remains a significant gap in formalizing the recommendations within frameworks and by healthcare professionals. Moreover, limited evidence exists on the effective implementation of these proposals in real-world scenarios.

Keywords. Prescriptive analytics, personalized medicine, prescriptive methodology, data science.

1 Introduction

The concept of Smart Health, which integrates Information and Communication Technologies such as the Internet of Things (IoT), wireless technologies, cloud computing, big data, and robotic systems, is emerging as the next evolution in the healthcare industry (Javed et al., 2022). This significant shift influences not only the organizational aspects of healthcare but also transforms the way medicine is practiced. Artificial intelligence (AI) and its techniques have notably advanced every science and research domain, including healthcare (Aceto et al., 2020).

The increasing availability of both structured and unstructured health data presents health professionals with the opportunity to develop more comprehensive patient profiles, enhancing the prescription process. Nonetheless, the vast volume of data presents challenges, particularly in time-sensitive and critical decision-making scenarios.

Medical prescriptions, often based on limited patient data and the healthcare professional's experience (N. Mosavi and Santos, 2020) underscore the context for applying Machine Learning (ML) techniques (Butryn et al., 2021). These techniques process data and serve as descriptive and predictive support tools. Yet, the full potential of data analytics in healthcare remains largely untapped, with a rising interest in applying prescriptive analytics to medical prescriptions, extending beyond organizational and operational considerations.

Acknowledging this interest is essential, along with recognizing ML techniques' limitations and understanding healthcare professionals' reluctance to depend on automated tools for critical decisions. Critiques of related works have pointed out significant issues like the "black box" effect of Convolutional Neural Networks, the extensive annotation required for Computer Vision, and the centralized and limited scope of many training and evaluation datasets for data analytics tools (Hechi et al., 2021).

Despite these challenges, medical associations have acknowledged AI's potential benefits for the healthcare field, including epidemiological analysis, phenotyping and risk stratification, diagnostics, automated reporting, and economic advantages (Alfano et al., 2020; Musacchio et al., 2020). Consequently, there is a pressing need for an epistemological foundation in Data Science research within the healthcare sector to establish best practices for ML techniques and to encourage the development of reliable and transparent AI systems for critical decision-making.

This work aims to conduct a systematic review of literature on methodologies that utilize prescriptive analytics for personalized treatment in healthcare, especially for chronic disease treatments. We consulted several specialized databases and assessed the results using the Systematic Mapping Process (Petersen, 2019). Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2009) and recommendations from Critical Appraisal Tools (Munn et al., 2014).

2 Background

The healthcare domain poses significant computational challenges across all fields, from the complexity, quantity, and variety of data involved to the regulations set by each country on medical data storage, use, and treatment. Recently, there has been considerable growth in both structured and unstructured medical data, presenting a great opportunity to transform the healthcare paradigm (Dalli et al., 2022).

Personalized medicine is based on the belief that individuals' unique characteristics at molecular, physiological, environmental exposure, and behavioral levels require tailored interventions (Goetz and Schork, 2018). This approach aims to provide "the right patient with the right drug at the right dose at the right time" (Pirracchio et al., 2019).

While this is not a new concept, technological advances in wearables, storage, communication, and hardware capabilities have led researchers to suggest that this paradigm is now achievable (N. Mosavi and Santos, 2020), (Lepenioti et al., 2020), (El Morr and Ali-Hassan, 2019).

The optimal treatment decision for a patient was formalized as a function $d(x)$ that maps a set of covariates, called X , to a treatment indicator $\{0, 1\}$, if a patient with $X = x$ receives treatment 1 then $d(x)=1$ but if receives treatment 0 then $d(x) = 0$, the value of a treatment decision is the average outcome if the treatment were applied to the entire target population (Petkova et al., 2017).

The best treatment decision is the one that optimizes $d(x)$. Additionally, if $Y^*(0)$ and $Y^*(1)$ represent the "potential outcome" of the patient after receiving treatment 0 or 1, then the potential outcome of the optimal treatment for a patient with covariates X could be expressed as follows:

$$Y(d) = Y^*(1)d(X) + Y^*(0) [1-d(X)]. \quad (1)$$

Prescriptive analytics is presented as the way to achieve this paradigm due to their optimization nature (N. S. Mosavi and Santos, 2022b). This type of analytics sits at the pinnacle of the analytical process and utilizes insights obtained from predictive and descriptive analytics, in addition to machine learning algorithms, business rules, computer models, and operations across diverse datasets to answer the question: what should we do (Poornima and Pushpalatha, 2020)?

However, this is merely a step in the Data Science process; the entire process encompasses collection, storage, preprocessing, analysis, and visualization of data (Sivarajah et al., 2017).

Each stage presents its own challenges and research areas, but applying a prescriptive perspective over the model, rather than the programming design, could offer benefits such as easier updates, clearer implementations, and a variety of design options and modifications (Poornima and Pushpalatha, 2020).

Most existing research on data analysis in literature relates to descriptive and predictive analytics. Systematic literature reviews have been conducted to analyze techniques, algorithms, and application fields. The literature related to prescriptive analytics and its benefits to multiple fields has also been revised, with each source suggesting healthcare as a viable application field, ranging from organizational decisions to personalized treatment designs.

As can be inferred, obtaining robust predictors about which treatment will best serve the patient in the subsequent time interval requires extensive

Table 1. First selection criteria

Criterion	Proposed score			
	3	2	1	0
Search terms appear in title, abstract or keywords	Three sections	Two sections	One section	
The research field is medicine			Yes	No
The work theme is personalized treatments generation			Yes	No

amounts of observational data. There are three significant initiatives for data gathering: i2b2 (Informatics for Integrating Biology and the Bedside) (*i2b2: Informatics for Integrating Biology and the Bedside*, s/f, p. 2), MIMIC (Medical Information Mart for Intensive Care) (*The Medical Information Mart for Intensive Care*, s/f), and METRIC (Multidisciplinary Epidemiology and Translational Research in Intensive Care Data Mart) (*METRIC Lab - Overview*, s/f).

Each of these provides accessible databases containing information about Intensive Care Units' admissions and treatments in the USA.

Other countries have adopted the idea of a centralized medical database or have proposals in place to achieve this for data analysis and improved public health decisions (Hassan et al., 2021). This approach also necessitates new metrics focused on evaluating the effectiveness of Individual Treatment Regimes compared to non-personalized ones (Imai and Li, 2021).

To determine the current stage of the prescriptive approach in treatment design in healthcare, a Systematic Literature Review (SLR) was performed, with the methodology presented in the following section.

3 Method

To achieve the previously stated objective, a systematic review of the literature was conducted to identify, assess, and categorize proposed methodologies that involve a prescriptive approach. These methodologies were then analyzed to determine which steps of the Data Science process they cover, as well as the techniques involved. Finally, the results were

classified based on the methods used to obtain prescriptive analytics.

3.1 Protocol and Training

The research method was conducted using an adapted version of the systematic mapping process proposed by Peterson et al. (Petersen, 2019), as shown in Fig 1.

The screening method was conducted following the PRISMA methodology (Moher et al., 2009), and the data extraction and evaluation were carried out using the critical appraisal tools recommendations.

As this work is related to methodologies, the critical questions were stated in metrics and organized in a rubric form, as shown in section 3.5. These metrics were reviewed and approved by specialists in data mining, artificial intelligence, and healthcare.

The screening method was conducted following the PRISMA methodology, and the data extraction and evaluation were conducted using the recommendations of critical appraisal tools. As this work relates to methodologies, the critical questions were expressed in metrics and organized in a rubric form, as shown in section 3.5.

These metrics were reviewed and approved by specialists in data mining, artificial intelligence, and healthcare.

The first step was to define the research questions to delimit the research scope. The questions were formulated as follows:

Q1.- Which specific prescriptive methodologies have been developed for generating personalized treatments in healthcare in the past five years?

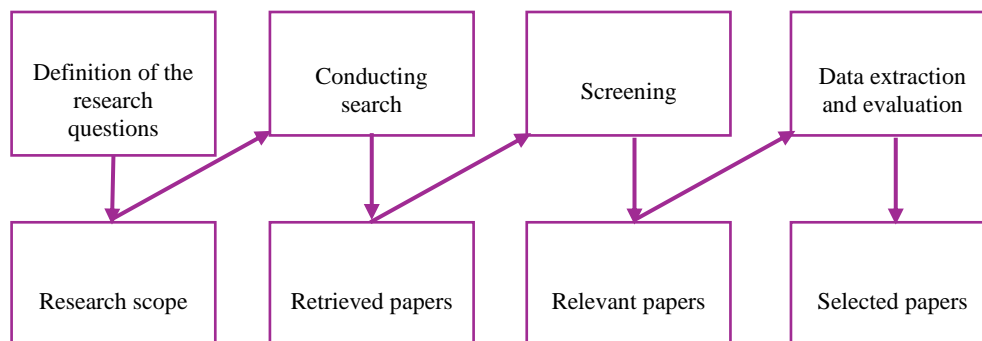


Fig. 1. Selection method. Source: Own elaboration

Q2.- How do these prescriptive methodologies integrate with the established data science lifecycle in healthcare analytics?

Q3.- What are the most effective techniques currently employed in prescriptive analytics for improving patient outcomes in healthcare?

The queried databases are presented in the next section.

4 Selected Databases and Queries

Following the recommendations of Martinovich (Martinovich, s/f), searches were conducted using a set of databases grouped as follows: aggregators (PubMed, Scopus, WOS, and BVS) and search engines (Google Scholar).

Search Methodology

The search terms employed were "prescriptive analytics" AND "medicine," "prescriptive methodology" AND "medicine," "prescriptive analysis" AND "medicine," and "prescriptive approach" AND "medicine." Subsequent searches replaced "medicine" with "healthcare."

These terms were utilized in both English and Spanish languages. The retrieved papers were organized using tools such as Zotero and Microsoft Excel to facilitate the screening process.

The selection criteria, as detailed in the following subsections, were applied during this phase. Each paper selected was confirmed independently by each reviewer.

Inclusion Criteria

The relevant set of papers included those that explicitly referred to the application of prescriptive analytics in the generation of treatments for healthcare-related diseases or ailments. Additionally, results proposing a methodology using a prescriptive approach, distinct from organizational operations in healthcare, were deemed eligible.

Prior surveys and reviews were also considered to ascertain their scope, metrics, and findings. However, conference papers and proposal work, often treated as "wish lists," were not included in the formal review.

Exclusion Criteria

The focus being specifically on the application of prescriptive analytics, works solely addressing descriptive and/or predictive analytics approaches were excluded.

These exclusion criteria also applied to editorial articles, book chapters, and works published in non-indexed journals.

Proposed Metrics

The initial paper selection was quantitatively evaluated using criteria presented in Table 1. This evaluation utilized a text analysis tool, which searched for the specified terms within three

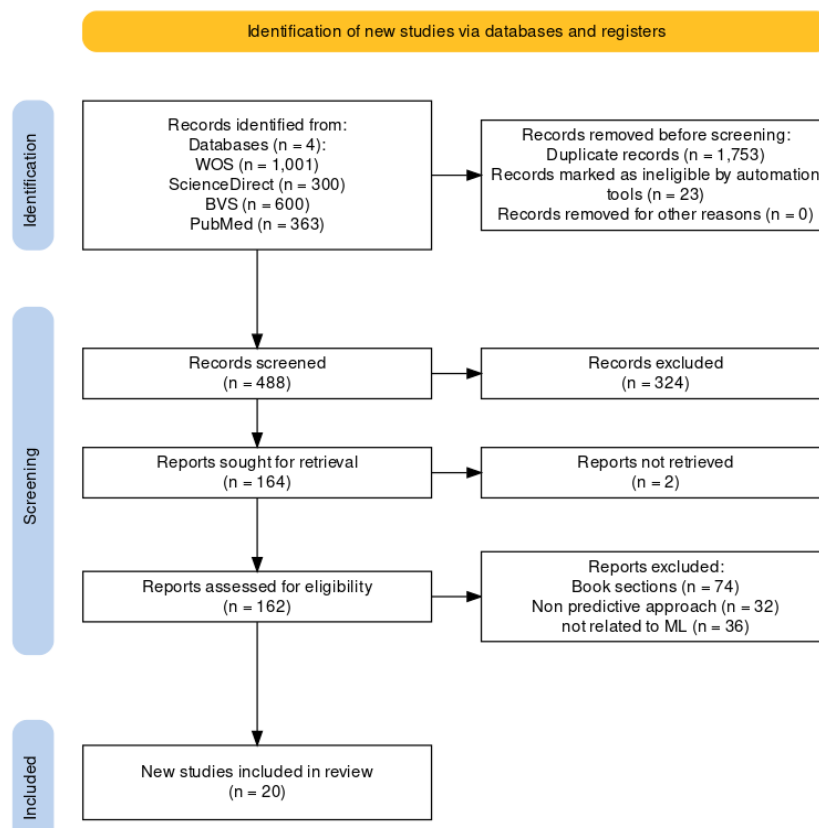


Fig. 2. Process of paper selection. Source: Own elaboration using PRISMA tool

sections of each paper: the title, abstract, and keywords.

The tool then returned the number of sections in which a match was found.

A score ranging from four to five points indicated that the work required a full-text review, while a score from one to three signified that the work needed corroboration from other reviewers to be considered. If a work was assigned a zero in either the second or the third criterion, it was rejected. Rejections at this stage were confirmed by humans.

A second quantitative evaluation was carried out based on the type of result following the full-text revision. If the work was a survey or review, it was evaluated using criteria shown in Table A1; for other types of papers, criteria in Table A2 were applied. These tables are presented in the annexes.

The application of these criteria led to a quantitative evaluation of relevant papers. For surveys and reviews, any paper with a score greater than or equal to 18 was included. For other types of papers, any with a score greater than or equal to 37 was included in the selected set of papers. This threshold represented 75% of the maximum score, established based on the reviewers' criteria.

5 Results

Following the established protocol, the process of constructing the selected set of papers was initiated. The prescribed queries were executed, leading to the retrieval of a specific set of papers.

A meticulous screening process ensued, incorporating the metric for the initial selection to

Table 2. Selected reviews and surveys

Reference	Reviewed works	Related area	Final product	Type of conclusions
(Thapa and Camtepe, 2021)	Not specified	Data regulation and security	None	Forward-looking
(Behera et al., 2019)	32	Cognitive computing	None	Forward-looking
(Khalifa, 2018)	56	Healthcare	Classification based on analytics utility	Descriptive
(Islam et al., 2018)	117	Healthcare	None	Forward-looking
(Baron, 2021)	Not specified	Healthcare	Example using queue mining	Descriptive
(Denton, s/f)	Not specified	Personalized medicine	Classification of challenges	Forward-looking
(Mehta et al., 2019)	2421	Healthcare	Map	Forward-looking
(Aceto et al., 2020)	171	Healthcare	Classification of challenges in three pillars	Descriptive
(Li et al., 2023)	Not specified	Healthcare	Classification of techniques	Descriptive
(Srivani et al., 2023)	75	Cognitive computing	Application recommendations	Descriptive

discern the collection of pertinent papers. Subsequently, each researcher independently scrutinized this set, engaging in a consensus-driven approach to curate the final selection of papers for further analysis. Fig 2 illustrates this process, from which 30 papers were obtained. Selected papers were categorized into two groups: (a) reviews and surveys, and (b) research works. A subset of the findings is displayed in Table 2, which outlines the selected surveys and systematic literature reviews (SLRs).

The column labelled "Related Area" identifies whether the review encompasses results from the entire healthcare domain or if it is focused on specific sub-areas. Additionally, the "Type of Conclusion" column denotes the nature of conclusions drawn in the scrutinized works.

The term "descriptive" indicates that authors summarize their findings, while "forward-looking" suggests they propose application guidelines or highlight future research directions.

Based on the contributions of the work, the scientific articles analyzed are categorized as shown in Fig 3: Within the "Method Proposal" category, these papers are focused on introducing

innovative algorithms for the generation of prescriptions.

In the "Case Study" classification, emphasis is placed on papers that utilize established ML methods to formulate prescriptive analytics for specific health conditions, thereby deviating from the primary focus on medical treatment prescriptions. The "Predictor Proposal" section includes papers dedicated to suggesting metrics for assessing prescriptive analytics.

The distribution of the Data Science phases involved in the selected works is presented in Fig 4. The discussion of the reviewed works is presented in the following section.

6 Discussion

The organization of this section is as follows: initially, the findings from each group of works are described; subsequently, specific aspects of some of the works are commented on; and finally, a general discussion is presented where the research questions are addressed.

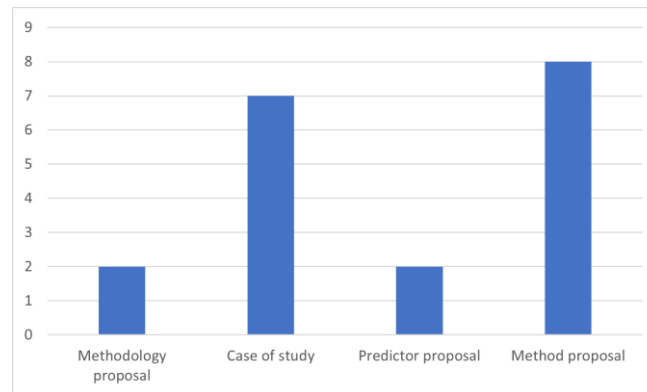


Fig. 3. Selected works by type. Source: Own elaboration

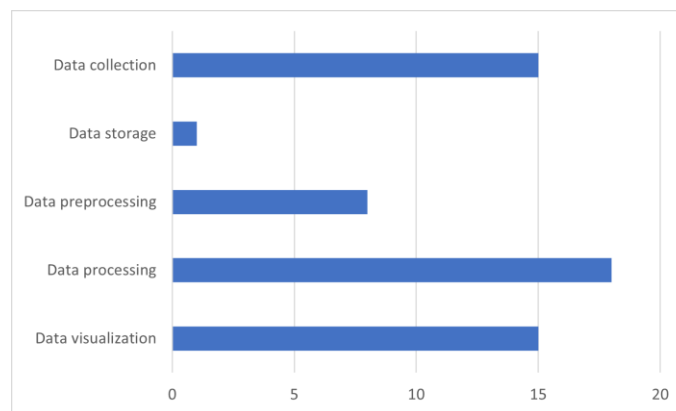


Fig. 4. Selected works by phases involved. Source: Own elaboration

Surveys and Reviews

Most of the studies identified in the literature are related to predictive and descriptive analytics, with only a limited number pertaining to prescriptive analytics. This observation aligns with findings reported by other reviewers (Islam et al., 2018). A recurrent theme across all reviewed works is the emphasis on personalized care as an emerging field of application, highlighting the interdisciplinary nature of these studies.

Various researchers have proposed different classifications in their systematic literature reviews (SLRs) (Baron, 2021; Islam et al., 2018; Khalifa, 2018). One classification is based on the utility of different types of analytics: descriptive, predictive, and prescriptive.

However, the classification of analytics remains an open point within the Data Science process, as

some works also mention other types, such as diagnostic and discovery analytics (Khalifa, 2018) or comparative analytics (Baron, 2021).

Another classification focuses on how analytics can enhance the organizational decision-making process, covering areas such as scheduling policies, resource matching, service paths, and diagnostics (Baron, 2021; Islam et al., 2018). Further proposals for classification are based on the applications of analytics within healthcare fields such as public health, administration, or mental health (Islam et al., 2018). The studies retrieved for analysis demonstrate a diverse range of considerations.

Some studies do not focus on any particular approach (Islam et al., 2018; Mehta et al., 2019), while others concentrate on specific methodologies or subareas (Li et al., 2023), or encompass all retrievable content (Baron, 2021;

Islam et al., 2018). Moreover, the strategies for advancing towards the development of prescriptive analytics vary. One strategy involves applying optimization techniques, such as data mining, to predictive models generated by any method (Baron, 2021). Another strategy suggests that a process incorporating a prescriptive analytics perspective should be adopted from the outset (Islam et al., 2018).

Regarding the types of works, the majority focus on the development and validation of algorithms, followed by topics such as evaluation metrics, governance, and data sources (Aceto et al., 2020; Denton, s/f; Mehta et al., 2019; Thapa and Camtepe, 2021).

The primary application areas for analytics are in oncology, neurology, and cardiology (Mehta et al., 2019); however, more recent works explore the application of methods to other conditions such as diabetes and dengue (Hoyos et al., 2021; Meng et al., 2020; Zheng et al., 2021).

There is a discrepancy in the scope of analytics considered in these studies, ranging from thousands of works analyzed encompassing all types of analytics (Mehta et al., 2019), to hundreds of works considering additional areas beyond healthcare (Aceto et al., 2020), to just a few specifically focusing on prescriptive analytics (Denton, s/f).

Among these studies, the need to incorporate contemporary technologies such as the Internet of Things, big data, and fog and cloud computing is recognized as crucial for enriching the context for decision-making processes (Aceto et al., 2020).

Additionally, the absence of existing regulations bridging the gap between regulation and innovation in the context of medical data is highlighted (Aceto et al., 2020; Denton, s/f; Thapa and Camtepe, 2021). Furthermore, the importance of deriving analytics from observational and longitudinal data and addressing bias is emphasized (Denton, s/f).

Research and Journal Works

The term "prescriptive methodology" was found, up to the point of the systematic literature review (SLR), in two significant works (Bertsimas et al., 2020; Hoyos et al., 2021). Bertsimas et al. propose a prescriptive methodology for the personalized

treatment of coronary artery disease patients using data from Electronic Medical Records (EMR).

They introduce an algorithm called ML4CAD, which integrates various machine learning predictive models through a voting mechanism, trained on data from 21,460 patients.

The effectiveness of the algorithm is evaluated using two novel metrics they propose: prescriptive effectiveness (PE) and prescriptive robustness (PR). Experimental results show a 24.11% improvement in potential adverse effect time compared to baseline treatments (Bertsimas et al., 2020).

To address missing values in 7,962 records, the authors generated artificial data based on each patient's context derived from EMR data. For descriptive analytics, k-NN clustering was performed to classify Time to Adverse Events (TAE) within 2-, 5-, and 10-year periods. For predictive analysis, five different tree-based methods were applied, including logistic regression, random forest, boosted trees, CART, and Optimal Classification Trees (OCT), using 31 characteristics. The data was randomly split to evaluate sensitivity.

The authors acknowledge several limitations of their methodology, including its non-multicentric nature, non-randomization of patients, omission of socioeconomic factors and patient preferences, and the non-representative nature of their sample. Moreover, they report that their results often did not align with standard care. An extension of this work applying a similar methodology to longitudinal hypertensive treatments was tested, but with the same limitations acknowledged by the authors (Bertsimas et al., 2022).

An additional methodology is proposed by Hoyos et al., where fuzzy cognitive maps (FCM) and genetic algorithms are combined to generate a prescriptive model and its inherent predictive model. Initially, an FCM is created from concepts in the domain model, followed using a genetic algorithm for decision optimization.

To test their proposal, the authors conducted experiments on three application cases: dose estimation, treatment selection, and contagion prevention, utilizing patient context information that includes sociodemographic, clinical, genetic, and laboratory data (Hoyos et al., 2021).

Table A1. Proposed criteria for surveys and reviews

Criterion	Proposed score			
	3	2	1	0
Number of related works	More than 20	From 10 to 20	Less than 10	
Time frame		Last five years	More than last five years	
Databases queried		More than three	Less than three	
Is the theoretical background based?			Yes	No
Is there a comparative analysis of the referenced works?		Present	Absent	
Is the query method reproducible?			Yes	No
The referenced works are present in the discussion section		All of them	More than a half of them	Less than a half of them
Evaluation metrics			Present	Absent
Type of conclusions		Prospective	Descriptive	
Are the results presented in a taxonomy or a classification?			Yes	No
Is the dataset available?			Yes	No

Table A2. Proposed criteria for research and proposals

Criterion	Proposed score			
	3	2	1	0
Type of work	Case study	Proposal	Other	
Type of research	Mixed	Quantitative	Qualitative	Unspecified
Type of publication		Journal paper	Conference paper	
Type of experiment	RCT	Cohort	Cases and controls	Unspecified
Experiment details			Specified	Unspecified
Based theoretical background			Present	Absent
State-of-art			Present	Absent
Data collected	Clinical Records	Public Dataset	Survey autogenerated	/ Unspecified
Data preprocessing			Detailed	Undetailed
Detail level of preprocessing	Complete	Partial	Commented	Unspecified
Analytics obtained	Prescriptive	Predictive	Descriptive	Unspecified
Method used for analytic generation		Explained	Commented	Unspecified
Additional methods			Present	Absent
Detail level of results			Complete	Partial
Results comparison			Present	Absent
Number of works for results comparison		Three or more	One or two	None

Comparison type	Quantitative	Qualitative	Unspecified
Comparison metrics		Present	Absent
Justification of the applied methods	Present detailed	and	Present but not detailed
Justification of the selected works for result comparison	Present detailed	and	Present but not detailed
Justification of the obtained results	Present detailed	and	Present but not detailed
Conclusions	Prospective	Descriptive	
Conclusions quality		Result-based	Not result-based
Dataset		Present	Absent
Source code		Present	Absent

Works proposing new algorithms for obtaining prescriptive analytics employ techniques such as Reinforcement Learning (RL). (Ahmed et al., 2021; Bertsimas et al., 2020, 2022; Laber and Staicu, 2017; Saghafian, 2023; Sun et al., 2023; Wang et al., 2019; Zheng et al., 2021), decision trees (Bertsimas et al., 2020, 2020; Laber and Staicu, 2017), Markov chains (Dasari et al., 2021; Meng et al., 2020), Bayesian inference (Rodriguez Duque et al., 2023), genetic algorithms (Hoyos et al., 2021) and mathematical programming (Kaur et al., 2018; Kessler et al., 2019; Moreno-Fergusson et al., 2021; Raychaudhuri et al., 2021).

Another perspective is divided on the use of data augmentation techniques from observational data (Kessler et al., 2019; Rose et al., 2023). However, the utilization of retrospective data for generating prescriptive analytics is mentioned as a limitation in several works (Bertsimas et al., 2020, 2022; Dasari et al., 2021; Meng et al., 2020).

Data used for the creation and training of models primarily focuses on clinical records (both electronic and non-electronic) (Ahmed et al., 2021; Laber and Staicu, 2017; Meng et al., 2020; Saghafian, 2023; Zheng et al., 2021), longitudinal patient data (Kessler et al., 2019; Moreno-Fergusson et al., 2021), observational information (Speth and Wang, 2021), real-time data (Raychaudhuri et al., 2021), and retrospective studies (Dasari et al., 2021).

Additionally, the significance of local context stands out as a crucial consideration; results obtained by tools constructed during the COVID-19 pandemic, which were designed and trained on data from a specific geographical region, seem to

display notable differences when utilized in a disparate region. This could suggest that context is relevant for treatment generation in the healthcare field (Bertsimas et al., 2020).

Proposed Frameworks

These works, while not directly related to treatment generation, are considered important as they set criteria for AI tools designed to support healthcare professionals in their decision-making processes.

Proposals were identified, crafted from the standpoint of healthcare professionals, aiming to establish a framework for developing health information management applications.

Among the recommendations, criteria such as scalability and the adoption of standards for health information management (HL7, FHIR, etc.) (Kaur et al., 2018), an uncertainty metric (Pirracchio et al., 2019), the use of the Internet of Things, wearables, and a well-defined layered structure that facilitates the identification of technology application levels within the tools (Kaur et al., 2018; N. Mosavi and Santos, 2022a) were emphasized.

Other aspects explored include real-time data access, data availability, multicentric and historical information, and the aggregation of information from multiple data sources, such as social networks (N. Mosavi and Santos, 2022a). Lastly, one study suggests that the limitations of the traditional data science approach could be mitigated by transitioning the focus from prediction to suggestion, emphasizing causal inference as a methodological approach (Pirracchio et al., 2019).

The scope of this SLR was outlined based on three questions provided earlier. Here, the responses to those questions are presented.

Q1. Which specific prescriptive methodologies have been developed for generating personalized treatments in healthcare in the past five years?

Two works proposing a prescriptive methodology were identified, although their dependence on available data introduces certain limitations. Acknowledging the constraints highlighted by the authors and the absence of information regarding the data storage phase, their incompleteness is evident.

Nevertheless, they signify notable advancements, showcasing the potential contributions of prescriptive analytics to the healthcare treatment domain.

Importantly, at the time of this SLR, these works represent instances specifically addressing the personalized generation of treatment and delineating the associated challenges for obtaining meaningful results.

The existing theoretical frameworks, while present, lack the granularity required to be considered methodological proposals, providing only a macroscopic view of the Data Science phases and their deployment using a prescriptive approach.

Q2. How do these prescriptive methodologies integrate with the established data science lifecycle in healthcare analytics?

The emphasis across all works lies on the data processing phase. However, only three of them present method proposals specifically for this phase, lacking information about the entire Data Science process.

Most theoretical frameworks assert that all phases should contribute to the objective of a prescriptive approach. Furthermore, nearly all the reviewed works lack details on how predictive analytics were obtained. Additionally, there is a dearth of information concerning the data storage phase, which could be relevant according to theoretical frameworks.

Q3. What are the most effective techniques currently employed in prescriptive analytics for improving patient outcomes in healthcare?

Overall, the reviewed works suggest that there is a need for a prescriptive approach in healthcare to move beyond predictive models and towards generating personalized treatment recommendations. However, there is a lack of well-defined methodologies for implementing such an approach, and most works focus on the analysis phase without providing sufficient information on other phases of the Data Science process. Different ML techniques have been proposed for generating prescriptive analytics, such as RL, MDP, decision trees, and simulations. Some works propose a combination of multiple techniques to increase the robustness of the recommendations.

Nevertheless, there is a need for greater transparency in the process of generating recommendations, as well as the inclusion of metrics to help clinicians in the decision-making process. Among the main problems mentioned by the works, the difficulty of consolidating a multicenter dataset is found due to various reasons: the sensitive nature of health information, the diversity of predictors used in treatment selection, and the heterogeneity of health information.

In addition, public datasets do not have a sufficient level of detail to build reliable models that can be taken to a higher level of study. Researchers propose techniques such as data augmentation, inference, "what-if" scenarios, reinforcement learning, and their variants to deal with these situations, but as seen, this is still in an initial phase.

Another problem is the lack of reliable metrics to evaluate the quality of a treatment suggestion; in medical prescription: What does optimal mean? This question needs to be answered by health professionals.

7 Conclusions

The expectations regarding the role of AI in the healthcare industry are ambitious, and Data Science researchers are motivated to contribute to achieving this goal. Theoretical works trust in prescriptive analytics obtained via Big Data as the guiding path for research. Although each ML technique has its pros and cons, the field of treatment generation is too crucial to rely on

common metrics. A multidisciplinary and multifactorial view of the problem is needed, based on the context of each patient's reality.

Each DS phase should be focused on a prescriptive approach, including the required tools for gathering data, the infrastructure and legal framework for storing data, the computational capacity to process all the information to obtain useful insights, and finally, a clear and intuitive visualization of results for health professionals to take care of the suggestions made by them.

This study was undertaken to evaluate diverse methodologies for prescriptive approaches. It is noteworthy that some excluded works put forth predictive methodologies and recommended a shift towards obtaining prescriptive analytics as future work. The findings indicate that there is potential for conducting an SLR for each phase of the Data Science process.

Moreover, the taxonomy of prescriptive methods proposed by Lepenioti et al. (El Morr and Ali-Hassan, 2019) can be leveraged to conduct an SLR focusing on techniques for constructing prescriptive analytics, thereby providing a more comprehensive response to the third research question.

References

1. **Aceto, G., Persico, V., Pescapé, A. (2020).** Industry 4.0 and health: Internet of things, big data, and cloud computing for healthcare 4.0. *Journal of Industrial Information Integration*, Vol. 18, DOI: 10.1016/j.jii.2020.100129.
2. **Ahmed, I., Ahmad, M., Jeon, G., Piccialli, F. (2021).** A framework for pandemic prediction using big data analytics. *Big Data Research*, Vol. 25. DOI: 10.1016/j.bdr.2021.100190.
3. **Alfano, C. M., Mayer, D. K., Beckjord, E., Ahern, D. K., Galioto, M., Sheldon, L. K., Hesse, B. W. (2020).** Mending disconnects in cancer care: setting an agenda for research, practice, and policy. *JCO Clinical Cancer Informatics*, Vol. 4, pp. 539–546. DOI: 10.1200/CCI.20.00046.
4. **Baron, O. (2021).** Business analytics in service operations—Lessons from healthcare operations. *Naval Research Logistics (NRL)*, Vol. 68, No. 5, pp. 517–533. DOI: 10.1002/nav.22011.
5. **Behera, R. K., Bala, P. K., Dhir, A. (2019).** The emerging role of cognitive computing in healthcare: a systematic literature review. *International journal of medical informatics*, Vol. 129, pp. 154–166. DOI: 10.1016/j.ijmedinf.2019.04.024.
6. **Bertsimas, D., Borenstein, A. R. A., Dauvin, A., Orfanoudaki, A. (2022).** Ensemble machine learning for personalized antihypertensive treatment. *Naval Research Logistics (NRL)*, Vol. 69, No. 5, pp. 669–688. DOI: 10.1002/nav.22040.
7. **Bertsimas, D., Boussioux, L., Cory-Wright, R., Delarue, A., Digalakis, V., Jacquillat, A., Zeng, C. (2021).** From predictions to prescriptions: A data-driven response to COVID-19. *Health care management science*, Vol. 24, pp. 253–272. DOI: 10.1007/s10729-020-09542-0.
8. **Bertsimas, D., Orfanoudaki, A., Weiner, R. B. (2020).** Personalized treatment for coronary artery disease patients: a machine learning approach. *Health Care Management Science*, Vol. 23, No. 4, pp. 482–506. DOI: 10.1007/s10729-020-09522-4.
9. **Butryn, B., Chomiak-Orsa, I., Hauke, K., Pondel, M., Siennicka, A. (2021).** Application of machine learning in medical data analysis illustrated with an example of association rules. *Procedia Computer Science*, Vol. 192, pp. 3134–3143. DOI: 10.1016/j.procs.2021.09.086.
10. **Dalli, J., Gomez, E. A., Jouvène, C. C. (2022).** Utility of the specialized pro-resolving mediators as diagnostic and prognostic biomarkers in disease. *Biomolecules*, Vol. 12, No. 3. DOI: 10.3390/biom12030353.
11. **Dasari, M., Perkins, J. D., Hendele, J. B., Leca, N., Biggins, S. W., Sibulesky, L. (2021).** Prescriptive analytics determining which patients undergoing simultaneous liver-kidney transplant may benefit from high-risk organs. *Experimental and Clinical Transplantation: Official Journal of the Middle East Society for Organ Transplantation*, Vol. 19, No. 12, pp. 1303–1312. DOI: 10.6002/ect.2021.0330.

12. **Denton, B. T. (2023).** Frontiers of medical decision-making in the modern age of data analytics. *IISE Transactions*, Vol. 55, No. 1, pp. 94–105. DOI: 10.1080/24725854.2022.2092918.
13. **El-Morr, C., Ali-Hassan, H., El-Morr, C., Ali-Hassan, H. (2019).** Descriptive, predictive, and prescriptive analytics. *Analytics in healthcare: a practical introduction*, pp. 31–55. DOI: 10.1007/978-3-030-04506-7_3.
14. **Goetz, L. H., Schork, N. J. (2018).** Personalized medicine: motivation, challenges, and progress. *Fertility and sterility*, Vol. 109, No. 6, pp. 952–963. DOI: 10.1016/j.fertnstert.2018.05.006.
15. **Hassan, S., Dhali, M., Zaman, F., Tanveer, M. (2021).** Big data and predictive analytics in healthcare in Bangladesh: regulatory challenges. *Heliyon*, Vol. 7, No. 6. DOI: 10.1016/j.heliyon.2021.e07179.
16. **El-Hechi, M., Ward, T. M., An, G. C., Maurer, L. R., El-Moheb, M., Tsoulfas, G., Kaafarani, H. M. (2021).** Artificial intelligence, machine learning, and surgical science: reality versus hype. *Journal of surgical research*, Vol. 264, pp. A1–A9. DOI: 10.1016/j.jss.2021.01.046.
17. **Hoyos, W., Aguilar, J., Toro, M. (2021).** Dengue models based on machine learning techniques: A systematic literature review. *Artificial Intelligence in Medicine*, Vol. 119, DOI: 10.1016/j.artmed.2021.102157.
18. **A National Center for Biomedical Computing (2022).** i2b2: Informatics for integrating biology and the bedside. <https://www.i2b2.org/>.
19. **Imai, K., Li, M. L. (2023).** Experimental evaluation of individualized treatment rules. *Journal of the American Statistical Association*, Vol. 118, No. 541, pp. 242–256. DOI: 10.1080/01621459.2021.1923511.
20. **Islam, M. S., Hasan, M. M., Wang, X., Germack, H. D., Md-Alam, N. E. (2018).** A systematic review on healthcare analytics: application and theoretical perspective of data mining. *Healthcare. Multidisciplinary Digital Publishing Institute*, Vol. 6, No. 2, p. 54. DOI: 10.3390/healthcare6020054.
21. **Javed, A. R., Shahzad, F., ur-Rehman, S., Zikria, Y. B., Razzak, I., Jalil, Z., Xu, G. (2022).** Future smart cities: Requirements, emerging technologies, applications, challenges, and future aspects. *Cities*, Vol. 129. DOI: 10.1016/j.cities.2022.103794.
22. **Kaur, P., Sharma, M., Mittal, M. (2018).** Big data and machine learning based secure healthcare framework. *Procedia computer science*, Vol. 132, pp. 1049–1059. DOI: 10.1016/j.procs.2018.05.020.
23. **Kessler, R. C., Bossarte, R. M., Luedtke, A., Zaslavsky, A. M., Zubizarreta, J. R. (2019).** Machine learning methods for developing precision treatment rules with observational data. *Behaviour Research and Therapy*, Vol. 120, DOI: 10.1016/j.brat.2019.103412.
24. **Khalifa, M. (2018).** Health analytics types, functions and levels: a review of literature. *Data, Informatics and Technology: An Inspiration for Improved Healthcare*, pp. 137–140. DOI: 10.3233/978-1-61499-880-8-137.
25. **Laber, E. B., Staicu, A. M. (2018).** Functional feature construction for individualized treatment regimes. *Journal of the American Statistical Association*, Vol. 113, No. 523, pp. 1219–1227. DOI: 10.1080/01621459.2017.1321545.
26. **Lepenioti, K., Bousdekis, A., Apostolou, D., Mentzas, G. (2020).** Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, Vol. 50, pp. 57–70. DOI: 10.1016/j.ijinfomgt.2019.04.003.
27. **Li, Z., Chen, J., Laber, E., Liu, F., Baumgartner, R. (2023).** Optimal treatment regimes: A review and empirical comparison. *International Statistical Review*, Vol. 91, No. 3, pp. 427–463. DOI: 10.1111/insr.12536.
28. **Martinovich, V. (2022).** Búsqueda bibliográfica: Cómo repensar las formas de buscar, recopilar y analizar la producción científica escrita. *Universidad Nacional de Lanús, Secretaría de Investigación y Posgrado*.
29. **Mehta, N., Pandit, A., Shukla, S. (2019).** Transforming healthcare with big data analytics and artificial intelligence: A

- systematic mapping study. *Journal of biomedical informatics*, Vol. 100. DOI: 10.1016/j.jbi.2019.103311.
30. **Meng, F., Sun, Y., Heng, B. H., Leow, M. K. S. (2020).** Analysis via Markov decision process to evaluate glycemic control strategies of a large retrospective cohort with type 2 diabetes: the ameliorate study. *Acta Diabetologica*, Vol. 57, pp. 827–834. DOI: 10.1007/s00592-020-01492-x.
 31. **METRIC Lab (2022).** Mayo Clinic. <https://www.mayo.edu/research/labs/metric/oerview>.
 32. **Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Prisma Group (2010).** Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International journal of surgery*, Vol. 8, No. 5, pp. 336–341. DOI: 10.1371/journal.pmed.1000097.
 33. **Moreno-Fergusson, M. E., Guerrero-Rueda, W. J., Ortiz-Basto, G. A., Arevalo-Sandoval, I. A. L., Sanchez-Herrera, B. (2021).** Analytics and lean health care to address nurse care management challenges for inpatients in emerging economies. *Journal of Nursing Scholarship*, Vol. 53, No. 6, pp. 803–814. DOI: 10.1111/jnu.12711.
 34. **Mosavi, N. S., Santos, M. F. (2022).** To what extent healthcare analytics influences decision making in precision medicine. *Procedia Computer Science*, Vol. 198, pp. 353–359. DOI: 10.1016/j.procs.2021.12.253.
 35. **Mosavi, N. S., Santos, M. F. (2020).** How prescriptive analytics influences decision making in precision medicine. *Procedia Computer Science*, Vol. 177, pp. 528–533. DOI: 10.1016/j.procs.2020.10.073.
 36. **Mosavi, N. S., Santos, M. F. (2022).** Internet of things for precision intensive medicine. *Procedia Computer Science*, Vol. 201, pp. 732–737. DOI: 10.1016/j.procs.2022.03.099.
 37. **Munn, Z., Moola, S., Riitano, D., Lisy, K. (2014).** The development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence. *International journal of health policy and management*, Vol. 3, No. 3, 123. DOI: 10.15171/ijhpm.2014.71.
 38. **Musacchio, N., Giancaterini, A., Guaita, G., Ozzello, A., Pellegrini, M. A., Ponzani, P., de-Micheli, A. (2020).** Artificial intelligence and big data in diabetes care: a position statement of the Italian Association of Medical Diabetologists. *Journal of Medical Internet Research*, Vol. 22, No. 6, DOI: 10.2196/16922.
 39. **Petersen, B. T. (2019).** Other considerations: Perspective and obligations of our societies and governmental organizations. *Techniques in Gastrointestinal Endoscopy*, Vol. 21, No. 4, DOI: 10.1016/j.tgie.2019.04.007.
 40. **Petkova, E., Ogden, R. T., Tarpey, T., Ciarleglio, A., Jiang, B., Su, Z., Trivedi, M. H. (2017).** Statistical analysis plan for stage 1 EMBARC (Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care) study. *Contemporary clinical trials communications*, Vol. 6, pp. 22–30. DOI: 10.1016/j.conctc.2017.02.007.
 41. **Pirracchio, R., Cohen, M. J., Malenica, I., Cohen, J., Chambaz, A., Cannesson, M., Lee, C., Resche-Rigon, M., Hubbar, A. (2019).** Big data and targeted machine learning in action to assist medical decision in the ICU. *Anaesthesia Critical Care and Pain Medicine*, Vol. 38, No. 4, pp. 377–384. DOI: 10.1016/j.accpm.2018.09.008.
 42. **Poornima, S., Pushpalatha, M. (2020).** A survey on various applications of prescriptive analytics. *International Journal of Intelligent Networks*, Vol. 1, pp. 76–84. DOI: 10.1016/j.ijin.2020.07.001.
 43. **Raychaudhuri, S. J., Manjunath, S., Srinivasan, C. P., Swathi, N., Sushma, S., Nitin-Bhushan, K. N., Narendra-Babu, C. (2021).** Prescriptive analytics for impulsive behaviour prevention using real-time biometrics. *Progress in Artificial Intelligence*, Springer Nature, Vol. 10, No. 2, pp. 99–112. DOI: 10.1007/s13748-020-00229-9.
 44. **Rodriguez-Duque, D., Stephens, D. A., Moodie, E. E., Klein, M. B. (2023).** Semiparametric Bayesian inference for optimal dynamic treatment regimes via dynamic marginal structural models. *Biostatistics*, Vol. 24, No. 3, pp. 708–727. DOI: 10.1093/biostatistics/kxac007.

45. **Rose, E. J., Moodie, E. E., Shortreed, S. M. (2023).** Monte Carlo sensitivity analysis for unmeasured confounding in dynamic treatment regimes. *Biometrical Journal*, Vol. 65, No. 5. DOI: 10.1002/bimj.202100359.
46. **Saghafian, S. (2024).** Ambiguous dynamic treatment regimes: A reinforcement learning approach. *Management Science*, Vol. 70, No. 9, pp. 5667–5690. DOI: 10.1287/mnsc.2022.00883.
47. **Sivarajah, U., Kamal, M. M., Irani, Z., Weerakkody, V. (2017).** Critical analysis of big data challenges and analytical methods. *Journal of business research*, Vol. 70, pp. 263–286. DOI: 10.1016/j.jbusres.2016.08.001.
48. **Speth, K., Wang, L. (2021).** Restricted sub-tree learning to estimate an optimal dynamic treatment regime using observational data. *Statistics in Medicine*, Vol. 40, No. 26, pp. 5796–5812. DOI: 10.1002/sim.9155.
49. **Srivani, M., Murugappan, A., Mala, T. (2023).** Cognitive computing technological trends and future research directions in healthcare — A systematic literature review. *Artificial Intelligence in Medicine*, Vol. 138. DOI: 10.1016/j.artmed.2023.102513.
50. **Sun, Z., Dong, W., Li, H., Huang, Z. (2023).** Adversarial reinforcement learning for dynamic treatment regimes. *Journal of Biomedical Informatics*, Vol. 137. DOI: 10.1016/j.jbi.2022.104244.
51. **Thapa, C., Camtepe, S. (2021).** Precision health data: Requirements, challenges and existing techniques for data security and privacy. *Computers in biology and medicine*, Vol. 129, DOI: 10.1016/j.compbiomed.2020.104130.
52. **Rogers, P., Wang, D., Lu, Z. (2021).** Medical information mart for intensive care: A foundation for the fusion of artificial intelligence and real-world data. *Frontiers in Artificial Intelligence*, National Library of Medicine. DOI: 10.3389/frai.2021.691626.
53. **Wang, Z., Li, M., Wang, H., Jiang, H., Yao, Y., Zhang, H., Xin, J. (2019).** Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. *IEEE* Vol. 7, pp. 105146–105158. DOI: 10.1109/ACCESS.2019.2892795.
54. **Zheng, H., Ryzhov, I. O., Xie, W., Zhong, J. (2021).** Personalized multimorbidity management for patients with type 2 diabetes using reinforcement learning of electronic health records. *Drugs*, Vol. 81, pp. 471–482. DOI: 10.1007/s40265-020-01435-4.

Article received on 18/05/2024; accepted on 17/07/2024.

**Corresponding author is Gustavo Emilio Mendoza-Olguín*