

Forecasting the Length-of-Stay of Pediatric Patients in Hospitals: A Scoping Review

Natalia B. Medeiros

Universidade Federal do Rio Grande do Sul

Flavio S. Fogliatto (✉ ffogliatto@producao.ufrgs.br)

Universidade Federal do Rio Grande do Sul

Miriam K. Rocha

Universidade Federal Rural do Semi-Árido

Guilherme L. Tortorella

Universidade Federal de Santa Catarina

Research Article

Keywords: Hospital length-of-stay, Forecasting models, Pediatric patients, Neonatal patients, Scoping review

Posted Date: March 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-289075/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Healthcare management faces complex challenges in allocating hospital resources and predicting patients' length-of-stay (LOS) is critical in effectively managing those resources. This work aims to identify approaches used to forecast the LOS of Pediatric Patients in Hospitals (LOS-P), and patients' populations and environments used to develop the models.

Methods: Based on a scoping review, 28 studies were identified and analyzed. Methods reported in the literature were classified according to the stage in which they are used in the modeling process: (i) pre-processing of data, (ii) variable selection, and (iii) cross-validation.

Results: Forecasting models are most often applied to newborn patients and, consequently, in neonatal intensive care units. Regression analysis is the most widely used modeling approach; techniques associated with Machine Learning are still incipient and mostly used in emergency departments to model patients in specific situations.

Conclusions: The studies' main benefits include informing family members about the patient's expected discharge date and enabling hospital resources' allocation and planning. Main research gaps are associated with the lack of generalization of forecasting models and limited reported applicability in hospital management. This study also provides a practical guide to methods applied for LOS-P forecasting and a future research agenda.

Background

With the increasing demand for health services, hospitals worldwide are operating under pressure to increase patient care quality while ensuring organizational survival [1]. The hospital environment is complex and imposes several managerial challenges related to optimal utilization of limited resources and the constant need to improve efficiency and reduce patients' length-of-stay (LOS) [2]. Resource utilization planning requires predicting patients' LOS since longer times imply lower turnover and higher costs, affecting the quality of patient care and reducing the availability of services to the population [3].

Forecasting pediatric patients' LOS (LOS-P) in hospitals is particularly important due to this population's *characteristics*. Children are at high risk of mortality and constitute a complex population from the medical perspective [4]. In addition, pediatric departments constantly struggle with capacity and overcrowding restrictions [2], which could be avoided by predicting the use of hospital resources, and better dimensioning of care and hospitalization capacity [5].

Forecasting patients' LOS is a subject widely studied in the literature through different methods and applications. Some *studies* use simple methods such as linear regression analysis (e.g., [6]), while more current studies use artificial intelligence techniques based on machine learning and deep learning (e.g., [7]). Literature review studies devoted to forecasting patients' LOS in hospitals mostly analyze the adult population. Almashrafi et al. [8] and Peres et al. [9] conducted systematic literature reviews to find predictors of patients' LOS in hospitals that should be considered in the generation of forecasting models. Atashi et al. [10], Hussain and Dunn [11], Lu et al. [12], and Verburg et al. [13] focused their reviews on the quality of the models proposed in the literature aiming at establishing a benchmark. Seaton et al. [14] was the only literature review that analyzed the pediatric population, identifying important factors when predicting neonates' LOS in neonatal units.

In this article we present a scoping review of the literature on LOS-P forecasting, addressing the current lack of *studies* that analyze the existing theoretical framework on the topic and map the main approaches used in the generation of LOS-P forecasting models. We present the main modeling techniques, their benefits and limitations, the environments investigated, and the types of pediatric populations considered. We close the article presenting practical implications and directions for future research.

Methods

This research is a scoping study that follows the methodology proposed by Arksey and O'Malley [15], which comprises five steps: (i) identification of research questions, (ii) identification of relevant studies, (iii) selection of studies, (iv) mapping of the data, and (v) collection, summarization and reporting of the results. These steps are subsequently detailed. Scoping reviews provide a transparent and rigorous mapping of the research area, producing an accessible summary of the research results and indicating the existing gaps [15]. Scoping reviews are suitable for topics with scarcity of studies and whose body of knowledge has not been consolidated through a systematic literature review.

Identification of research questions

This research was driven by elements related to the forecasting of LOS-P, namely: techniques used, environments and populations in which techniques were applied, and results generated by the models created. Based on these elements, the study addressed five research questions:

RQ₁. What are the main techniques used to forecast LOS-P?

RQ₂. In which situations are those techniques applied?

RQ₃. What are the main characteristics of the data used to predict LOS-P?

RQ₄. What are the managerial implications of using LOS-P forecasting models for resource management in hospitals?

RQ₅. What are the main barriers and limitations of the existing studies and opportunities for future research?

Identification of relevant studies

The identification of relevant studies was structured around three axes related to the research questions: (i) forecast models, (ii) hospital length-of-stay, and (iii) pediatric patients. Those axes led to a combination of keywords used in the search (Table 1). Five databases were consulted: Pubmed, Science Direct, Scopus, Web of Science, and Medline. The first four bases were chosen following the recommendation of Tortorella et al. [16], who conducted a scoping study on a related topic. The inclusion of Medline was due to its recurrent use in systematic review articles on hospital LOS forecasting, e.g., [11], [14], [13]. The first axis should be present in the title, abstract, or keywords in our search, while the second and third axes should be present only in the title. The search was carried out between August and September 2020; no filters were applied, resulting in a total of 1,133 articles.

Selection of studies

To conduct this step, we followed the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) by Moher et al. [17]. The PRISMA guidelines are composed of four stages: identification, screening, eligibility, and inclusion (see Figure 1). The first two authors primarily carried out all stages to ensure the reliability of the review process. Differences in opinions were discussed with the remaining authors to reach a consensus.

Table 1 – Research axes and keywords

Research axes	Forecast models	Length of stay in hospital	Pediatric patients
Keywords	Predict*	"Length of Stay"	Child*
	Model	"Hospital Days"	Pediatric
	Prognos*	"Length of Hospital Stay"	Paediatric
	Forecast*	"Duration of Stay"	Kid
	Regression	"Patient Stay"	Youth*
	Estimat*		Adolescen*
			Neonat*
			Newborn*
		Infant*	

Note: * is used so that the search returns all terms that begin with the word followed by the asterisk. Enclosing a sentence in double-quotes ensures that the search returns only documents in which the sentence appears and not just the words in any order.

The identification stage was performed by searching for the axes in the chosen databases. The screening stage included three filters: (i) type of publication and language, (ii) exclusion of duplicates, and (iii) inclusion and exclusion criteria. The first filter selected journal or conferences articles published in English, eliminating, for example, abstracts from conferences and clinical tests; 81 articles were excluded, and 1,052 publications remained. The second filter excluded 759 duplicates, and 293 publications remained. In the final filter, the remaining articles' titles and abstracts were verified according to the established inclusion and exclusion criteria.

To be included in the review, studies should develop mathematical models to predict patients' LOS in hospitals, and the studied population should be pediatric patients. Publications were excluded when the dependent variable was not the LOS in hospitals (e.g., cost and efficiency) or investigated the effect of individual factors on LOS-P. After applying the inclusion and exclusion criteria, 56 articles remained. The eligibility stage consisted of a full-text analysis identifying articles aligned with the proposed research questions, resulting in 26 articles. In the final stage (inclusion), the selected articles' references were checked to identify studies potentially aligned with the research theme and not identified during searches in the selected databases. Based on this backward snowballing technique, two articles were included totaling 28 publications in the final *corpus*.

Mapping of the data

According to Arksey and O'Malley [15], data mapping must extract key information from the studies being reviewed to help readers make decisions. We used the descriptive-analytical method to build an analytical structure, collecting standard information about each study. The following information was retrieved from our *corpus*, allowing the mapping of the literature: authors, year of publication, country of origin of the first author, journal or conference title, area of knowledge of the publication, application focus (i.e., department and population), data characteristics, techniques used to build the forecast model, limitations, contributions, and future research directions.

Collection, summarization, and reporting of results

In this step, we used the information extracted during the data mapping to group, summarize, and present the results in a logical and organized way, answering the research questions. We prepared an overview of the selected studies based on a descriptive numerical summary and thematic analysis [15].

The descriptive numerical summary provided a quantitative view of the articles' main characteristics. The thematic analysis allowed a qualitative view of the articles and in-depth insights regarding the literature. That provided means to complement the descriptive numerical summary trends, contributing to fully answer our research questions.

Results

Our corpus of works on LOS–P forecasting contains mostly journals papers (89%) authored by 125 authors who contributed to the topic with only one publication. Studies appeared in the proceedings of 3 conferences and on 21 different journals, two of which with more than one publication: Journal of Perinatology ($n=2$) and Journal of the American Academy of Child & Adolescent Psychiatry ($n=2$). It is evident the lack of an outlet concentrating studies on the topic. Research took place in nine different countries, four of which with more than one publication: USA ($n=17$), Germany ($n=2$), Brazil ($n=2$), and Canada ($n=2$). Studies belong to several knowledge areas, with prevalence in Pediatrics ($n=8$) and Psychiatry ($n=4$), medical specialties (e.g., cardiology and neurology), and medical departments (e.g., emergency and intensive care). Only one study is in the area of Computer Science, indicating that research on LOS–P prediction emphasizes healthcare applications rather than forecasting methods. The evolution of publications per year shows an increase in studies over the decades, beginning in the 1980s ($n=1$) and with the most recent publications in 2020 ($n=2$). In the decades of 1990 ($n=6$), 2000 ($n=9$), and 2010 ($n=10$), the number of studies increased considerably, bending towards the use of Machine Learning modeling techniques.

In the thematic analysis, we divided studies according to three dimensions: the technical approach used to generate the forecasts, the medical department where the study took place, and the population analyzed. Table 2 summarizes our findings and serves as a guide to what has already been reported in the literature, addressing RQ_1 and RQ_2 .

Table 2 – Summary of studies regarding forecasting approach, and investigated department and population

Department	Population	Approach			# of articles
		<i>Regression</i>	<i>Machine Learning</i>	<i>Others</i>	
Emergency	Babies and children with bronchiolitis		[42]		2
	Pediatric trauma patients		[19]		
Neonatal Intensive Care units	Premature newborns	[38], [41], [32]			11
	Newborns	[26], [39], [29], [23]		[28]	
	Chronically underweight newborns	[36], [40], [18]			
Pediatric Intensive Care units	Pediatric patients	[27]			1
Pediatric unit or hospital	Babies undergoing bidirectional Glenn procedure	[21]			3
	Children with hematological diseases complicated by febrile neutropenia	[31]			
	Pediatric patients with respiratory diseases		[2]		
Psychiatric unit or hospital	Children	[30], [25], [3]		[20]	6
	Teenagers	[37], [3], [24]		[20]	
	Young Adults	[37]			
Not specified	Premature newborns	[35]	[35]		5
	Newborns and babies undergoing cardiac surgery	[33]			
	Babies admitted for gastroenteritis	[43]			
	Pediatric patients		[34]		
	Pediatric victims of ATV accidents	[22]			
Nº of articles		21	5	2	

References: [21] Anderson et al. (2009); [34] Balan et al. (2019); [36] Bannwart et al. (1999); [28] Bender et al. (2012); [37] Browning (1986); [30] Gold et al. (1993); [38] Hintz et al. (2009); [20] Höger et al. (2002); [41] Jeremic and Tan (2008); [25] Kavanaugh et al. (2019); [26] Khoshnood et al. (1996); [43] Lee et al. (2005); [39] Lee et al. (2016); [3] Leon et al. (2006); [27] Levin et al. (2012); [2] Ma et al. (2020); [40] Marshall et al. (2012); [22] Nagarsheth et al. (2011); [33] Parkman and Woods (2005); [31] Pastura et al. (2004); [32] Paul et al. (2020); [29] Pearlman et al. (1992); [23] Pepler et al. (2012); [18] Rendina (1998); [24] Stewart et al. (2013); [19] Walczak and Scorpio (2000); [42] Walsh et al. (2004); [35] Zernikow et al. (1999).

The technical approaches were divided into Regression Analysis (used in 75% of the studies), Machine Learning techniques, and Others. Departments where studies took place were divided into six categories, one of which dedicated to articles that did not convey that information. The largest number of studies took place in Neonatal Intensive Care units (39.29%) and Psychiatric units or hospitals (21.43%), which are highly controlled areas with abundant LOS data. In

those environments, regression analysis was the predominant forecasting technique. In opposition, LOS data from Emergency departments were exclusively modeled using Machine Learning techniques. Regarding the population analyzed, most studies used data from newborn patients (42.86%) or patients in specific situations (28.57%), e.g., victims of ATV accidents and children with hematologic malignancies complicated by febrile neutropenia. Fewer studies (32%) involved adolescent patients or young adults.

Table 3 characterizes the datasets used in the studies, addressing RQ_3 . The information presented includes country of origin, sampling period, sample size, number of hospitals included in the sample, mean or median of the LOS-P.

Hospitalizations of pediatric patients were sampled in 13 countries; the USA ($n=14$) is the country with the highest representation in the studies (50%). Data were collected between 1987 and 2017, covering from 9 to 120 months; data were mostly collected during the 1990s ($n=10$) and 2000s ($n=10$). Four studies do not specify the sampling period; the majority performed the sampling in a period equal to or greater than one year ($n=22$). Sample sizes range from 41 to 23,551 observations. Most studies ($n=15$) took place at a single location, indicating a low concentration of multicenter studies. Studies' LOS-P values display averages or medians ranging from 3.39 days to 18.02 months, with more than 40% of the articles not reporting this information ($n=12$). The longest LOS occurred in hospitals or Psychiatric units (ranging from 2 to 18 months), indicating more extended stays and lower turnover in those types of services. In Neonatal Intensive Care units, LOS-P averages vary from 23 to 54.8 days, with the highest averages (54.8 and 52.8 days) concentrated in the population of very low weight neonates.

Table 3 - Characteristics of datasets used in the studies

Reference	Country	Sampling period	Sample size	Hospitals	LOS-P
[21]	USA	July 2001 to December 2007	100	1	Median: 20 days
[34]	USA	2016	5,236	4,200	Not informed
[36]	Not informed	January 1992 to December 1993	97	1	Mean: 52.8 days
[28]	USA	August 1999 to October 1999, and April 2002 to September 2002	908	1	Not informed
[37]	Not informed	Not informed	41	1	Mean: 18.02 months
[30]	USA	May 1988 to December 1989	96	1	Mean: 71.6 days
[38]	USA	July 2002 to December 2005	2,254	Not informed	Not informed
[20]	Germany	Not informed	1,001	13	Median: 104 days
[41]	Canada	Not informed	186	1	Not informed
[25]	Not informed	2010 to 2015	96	1	Mean: 18.56 days
[26]	USA	1990	558	1	Mean: 23 days
[43]	Australia	1995	514	58	Mean: 3.39 days
[39]	USA	2008 to 2011	23,551	125	Not informed
[3]	USA	1998 to 2001	1,930	44	Mean: 10.4 days
[27]	USA	Not informed	2,062	1	Mean: 3.5 days
[2]	China	January 2014 to April 2016	11,206	1	Not informed
[40]	Argentina, Chile, Paraguay, Peru and Uruguay	January 2001 to December 2008	7,599	20	Not informed
[22]	USA	January 2000 to December 2009	420	Not informed	Not informed
[33]	Not informed	September 1993 to December 1997	458	1	Not informed
[31]	Brazil	February 2001 to May 2002	62	1	Mean: 10 days
[32]	USA	November 2014 to March 2017	152	14	Not informed
[29]	USA	October 1987 to July 1988	393	1	Mean: 23.8 days
[23]	South Africa	January 2007 to December 2008	3,794	15	Mean: 17.9 days
[18]	USA	January 1994 to December 1996	314	2	Mean: 54.8 days
[24]	Canada	October 2005 to March 2010	2,445	69	Mean: 16.31 days
[19]	USA	April 1994 to December 1997	7,665	Not informed	Mean: 3.98 days
[42]	Ireland	1999	119	1	Not informed
[35]	Not informed	October 1989 to January 1996	2,144	1	Not informed

Methods used to build the forecasting models are divided into three groups, according to the stage of model development they propose to address: (i) pre-processing, (ii) variable selection, and (iii) cross-validation. Pre-processing consists of preparing the data prior to modeling, avoiding noise due to outliers, missing data, multicollinearity, and lack of variable standardization. Variable selection methods focus on optimizing the forecasting model by improving its precision and interpretability using only the most informative variables. Cross-validation is used to evaluate the performance of the model in different datasets. Table 4 presents the studies' main methods and the results obtained from their application, addressing RQ_1 . The pre-processing methods reported in our *corpus* may be divided into (i) data cleaning methods to avoid modeling noise and (ii) methods to prepare and transform data to remove scale effects. The main pre-processing method aimed at data cleaning is the collinearity test, which evaluates the correlation level between independent variables. The test was reported in eight studies [18], [19], [20], [21], [22], [23], [24], [25]. To avoid noise in the model due to the excessive number of observations with missing data in the independent variables, six studies excluded incomplete observations from the datasets [26], [18], [19], [27], [2], and two adopted data imputation strategies [28], [2]. Two studies mention the withdrawal of outliers from the datasets before modeling [23], [25].

The main pre-processing method for data transformation is the logarithmic transformation of LOS-P values to correct the positive asymmetric distribution of the dependent variable, adopted in ten studies [29], [30], [26], [20], [31], [22], [23], [18], [24], [32]. The logarithmic transformation also reduces the effect of

outliers, ensuring the normality of the residuals and stabilizing the variance. Machine Learning-based approaches did not transform the dependent variable.

Table 4 – Forecasting modeling approaches

Reference	Pre-processing							Variable selection		
	Logarithmic transformation	Coding of categoric variables	Data rescaling	Colinearity test	Variable categorization	Missing data treatment	Feature engineering	Outliers withdrawn	Backward stepwise selection	Forward stepwise
[21]				X						
[34]			X							
[36]					X					X
[28]						X				
[37]										
[30]	X									
[38]									X	
[20]	X			X						
[41]										
[25]				X				X		
[26]	X	X				X				
[43]										
[39]									X	
[3]		X					X			
[27]						X				X
[2]		X				X	X			
[40]										
[22]	X			X						
[33]		X			X					
[31]	X				X					
[32]	X								X	
[29]	X	X								
[23]	X			X				X		
[18]	X	X		X		X				
[24]	X	X		X						
[19]		X		X		X	X			

[42]										
[35]			X							X
Frequence	10	8	2	8	3	6	3	2	3	3
*AMSE = Average MSE, *PP = Perfect prediction, *P1 = Predictions up to one day, *CPR = Correlation between the predicted and actual LSPPH										

A large number of studies use the coding of categorical variables ($n=8$) through the use of dummy variables [18], [3], [2], [24], and specific codings that vary according to needs [29], [26], [19], [33]. Two studies use data rescaling through normalization [34] and linear rescaling [35]. Other data preparation methods include the categorization of variables [36], [31], [33], and feature engineering [19], [3], [2], which creates new variables by combining the ones available in the dataset.

Regarding the variable selection methods, most studies ($n=17$) propose reducing the number of variables in the model to keep only the most informative ones. Variable selection aims to balance model simplicity and performance; however, it is noteworthy that most Machine Learning-based studies do not use any variable selection method (except for Zernikow et al. [35], who proposes the Forward Stepwise method). The most popular variable selection methods are the Analysis of Variance [37], [30], [36], [20], [31], [23], and the variable significance test [29], [33], [3], [38], [28]. Both methods are statistical-based and straightforward, aimed at verifying the relationship between the dependent and independent variables.

Stepwise variable selection methods (backward or forward) perform the selection in stages, with the addition or removal of variables and subsequent assessment of the model's performance at each iteration. Three studies used the stepwise backward method [38], [39], [32], which starts with all variables in the model and removes at each iteration the least significant one. Three studies used the stepwise forward method [36], [35], [27], starting with a model with no variables and adding the most significant one at each iteration. In addition to the methods mentioned above, others less predominant in the studies are correlation analysis [25], [32], stepwise multiple Cox regression [40], and Principal Component Analysis (PCA) [23].

Cross-validation allows measuring model stability. Half of the studies in our *corpus* used the procedure to validate model results, seeking its generalization. Cross-validation approaches are divided into three categories: traditional holdout [35], [20], [3], [41], [38], [27], [34], [2], temporal holdout [19], [23], [28], [39], [2], and k -fold [42], [28].

The holdout method divides the dataset into two partitions (training and testing), which are mutually exclusive and vary according to the analyst's preferences. Traditional holdout randomly splits the dataset assuming that the frequency distributions do not change over time; temporal holdout divides the dataset taking into account the temporal evolution of the data. Except for two studies that do not mention the percentage of the dataset used in each partition [40], [2], all other studies used traditional holdout, varying the proportion of the dataset in the training and testing partitions, as follows: 80% – 20% [27], [34], 75% – 25% [35], 70% – 30% [38], and 50% – 50% [20], [3]. The k -fold method randomly divides the dataset into k unique parts of the same size, trains the forecast model with parts, and uses the remaining part for validation. The process is repeated times, such that all parts are used in the validation step. Walsh et al. [42] use 5-fold, with the dataset divided into training, testing, and validation in five different ways, while Bender et al. [28] do not give details on the k -fold method used.

The performance of LOS–P forecasting models was measured using several metrics, with only one study not detailing the applied model's performance [43]. The most used performance metric ($n=19$) is the coefficient of determination (R^2) that measures the proportion of the variability in the dependent variable captured by the model. In studies measuring model performance using R^2 , indicated in the last column of Table 4, values ranged from 0.04 to 0.9415, with most studies reporting values greater than 0.5 ($n=8$). Other performance metrics reported were the Root Mean Square Error (RSME) and the Mean Absolute Error (MAE). RMSE measures the standard deviation of model errors; studies that used this metric reported values between 0.296 and 18.8 days [39], [2]. MAE measures the absolute average value of the differences between forecasts and actual observations; studies that used this metric reported values between 2.5 and 14.6 days [19], [39].

Performance metrics reported in only one study were Average Mean Squared Error – ASME [40], the proportion of forecasts with up to 12 hours [27] and 24 hours [19] of error with respect to actual values, average correct performance [42], correlation between forecasts and actual values [35], proportion of perfect forecasts [19], correlation between predictions based on variables available at birth and 30 days after birth [40], and variance representing the amount of information captured by each independent variable in the model [30].

All studies reported benefits from using LOS–P forecasts for the hospital ecosystem, except for eight studies [37], [30], [18], [20], [21], [22], [24], [34], that focused only on model performance. We identified five dimensions positively affected by the use of forecasting models, partially addressing RQ_4 : (i) patient care, (ii) costs, (iii) hospital management, (iv) quality measurement, and (v) updating of medical practices.

In the patient care dimension, the two main benefits reported are providing families information about the expected discharge date [29], [36], [35], [33], [43], [40], [28], [23], and preventing complications associated with prolonged hospitalizations [43], [40], [25]. Hintz et al. [38] suggested that LOS–P prediction allows a better understanding of risk factors associated with prolonged stays. Identifying such patients may direct hospitals towards more aggressive treatments and the provision of specialized care to prevent complications [43], [25].

Benefits associated with the cost dimension are estimates of financial values spent on hospitalization [43] and cost reduction for the hospital [36], [28], [26], [40], [32]. LOS–P forecast contributes to the hospital's strategic planning and guides medical care, reducing the length of the patient's stay and, consequently, hospitalization costs [36], [40].

Hospital management can bring several benefits to the hospital, being directly related to the other dimensions. Studies report management areas positively affected by LOS–P forecasts, such as resource allocation and planning [28], [39], [2], [23], [19], [35], patient flow management [27], [19], hospital bed management [43], [27], [2], optimization of decision making [41], [27], [2], [42], [35], and shift staff scheduling [28], [35]. The availability of LOS forecasts at patient's admission allows an efficient allocation of resources [39]; identifying LOS predictors may also contribute to the effective management of scarce medical resources [2].

As for the benefits associated with measuring patient care quality, studies advocate the monitoring of hospital performance [43], [31], [23], [35], and the standardization of care across hospitals [39], [3], [40], [35]. Hospital performance may be assessed by measuring the effect of hospital-related variables in the LOS–P model [43] or by using predicted LOS values as reference [23], and the difference between the expected and actual LOS values as a service quality indicator [35]. Pepler et al. [23] suggest that performance measurement based on LOS–P forecasts should be part of a quality program monitored by hospital managers to improve patient care quality. Leon et al. [3] argue that efforts to maintain quality should be directed towards understanding variations in practice standards across hospitals, while Lee et al. [39] suggest standardizing the treatment of premature babies among institutions. By comparing LOS–P values across centers, benchmarking analyses may be performed, contributing to hospitals' strategic planning [40].

The last dimension of benefits is related to the detection of variations in historical patterns due to changes in medical practices resulting from updating LOS–P models [42]. Walczak and Scorpio [19] report that the use of neural network models makes solutions non-static; as medical practices evolve, the prediction model is quickly adapted through continuous learning based on new datasets.

To start addressing RQ_5 we list the limitations and barriers reported in the LOS–P forecasting literature. They are: (i) lack of model generalizability due to differences across hospitals [29], [26], [36], [20], [31], [33], [3], [38], [28], [27], [40], (ii) lack of data on potentially useful LOS–P predictors [36], [31], [42], [33], [3], [40], [24], [39], [25], (iii) small sample sizes used to obtain forecasting models [36], [42], [31], [38], [22], [32], and (iv) studies based on samples that do not adequately represent the entire population [42], [3], [21], [25].

Other limitations cited by the investigated authors include the presence of missing data in the dependent variable [3], [33], imperfect data collection resulting in noisy samples [37], [28], [27], inconsistencies in parameters' estimations [2], lack of consensus regarding the minimum precision level that enables using the model to support decision making [27], presence of multicollinearity between independent variables [20], and poor model performance when predicting large LOS–P values [35].

Discussion

Our review on LOS–P forecasting found that the most studied department is Neonatal Intensive Care Units (39.29%), aligned with the conclusion in Seaton et al. [14], who also reviewed LOS–P forecasting of a pediatric population. We also found that most studies used data from newborn patients (42.86%) or patients in specific situations (28.57%), and few studies used data from the general pediatric population. Lu et al. [12] provided a systematic review of risk adjustment models for hospital LOS. They concluded that only 22% of the studies analyzed LOS of general populations, while 78% focused on populations with specific diseases.

Datasets used in our studies indicated that most of them took place at a single location, aligned with findings in review articles by Atashi et al. [10] and Almashrafi et al. [8]. In opposition, Seaton et al. [14] and Verburg et al. [13] reported that most LOS forecasting studies were multicenter in their review articles.

To provide a practical guide for LOS–P forecasting, we investigated the methods applied in the studies and found that the main technical approach used was regression analysis (75% of the studies). That is aligned with findings in Hussain and Dunn [11], who reviewed studies that predicted LOS of thermal burned patients and reported that all forecasting models were based on multivariate regression analysis. We also reported a trend towards artificial intelligence-based forecasting models in recent years, which was not observed in previous review studies, e.g. Seaton et al. [14] reported only one model based on Artificial Neural Networks, while Almashrafi et al. [8] and Lu et al. [12] reported a model based on decision trees. The dependent variable used in all studies we reviewed was the continuous LOS; in opposition, Almashrafi et al. [8], Atashi et al. [10], Lu et al. [12], and Seaton et al. [14] presented studies that modeled the discrete LOS of patients using multivariate logistic regression.

The main pre-processing methods reported in our review were LOS–P logarithmic transformation, multicollinearity testing, and coding of categorical variables. Regardless of the frequently observed positive asymmetric distribution of the LOS variable, only 35.7% of our reviewed studies used the logarithmic transformation, similar to what was reported by Lu et al. [12], in which 32% of studies used log-transformed LOS as the dependent variable. Most studies in our review did not mention the handling of missing data, a gap that was also reported by Atashi et al. [10].

Analysis of variance and significance tests were the preferred approach for variable selection in our studies. Reviews by Hussain and Dunn [11] and Peres et al. [9] reported that most studies selected variables to be included in the forecasting models through univariate analysis, while Seaton et al. [14] reported results similar to ours. Some variable selection techniques reported in our review (i.e., stepwise multiple Cox regression, correlation analysis and Principal Components Analysis) were not reported by other authors.

In our review, most authors used the traditional holdout with simple random sample split as a cross-validation method. Previous reviews reported similar results: Seaton et al. [14] concluded that over 50% of the authors investigated had validated results by splitting the sample, while Verburg et al. [13] reported that over 38% of the studies also used simple random sample split. Bootstrapping cross-validation, an approach not reported in our studies, was mentioned by Atashi et al. [10] and Verburg et al. [13].

Finally, regarding the performance of LOS–P forecasting models, several metrics were used in our reviewed studies, but the coefficient of determination (R^2) was predominant, with values that ranged from 0.04 to 0.9415. The same predominance was found in the studies reported by Seaton et al. [14] and Hussain and

Dunn [11], with values ranging from 0.158 to 0.75. In opposition, reviews by Atashi et al. [10] and Verburg et al. [13] found that Pearson's correlation coefficient was the preferred performance metric.

Conclusion

We close this paper further addressing RQ_4 and RQ_5 . In our study, we aimed at identifying studies devoted to LOS–P forecasting through a scoping review of the literature. LOS forecasting is a tool to improve the management of resources in healthcare environments, helping organizations to cope with high demands and quality requirements. Although some literature review papers investigate the prediction of LOS in different environments and patient populations, we bridge a gap in the literature regarding reviews focusing on pediatric patients, a complex population with a high mortality risk that represents a challenge for hospital managers.

This work has practical implications as it provides arguments to evaluate the applicability of different modeling techniques in forecasting the LOS–P in different environments and types of pediatric populations. That allows informing family members about the patient's expected discharge date, which is particularly critical for pediatric patients [44]. Additionally, hospital resources' allocation and planning can be significantly improved by properly estimating the LOS–P. More assertive and precise LOS–P forecast models can significantly avoid waste of labor and materials' shortages in hospitals, reducing costs and increasing the efficiency [9].

Levin et al. [27] cite that managing patients' flow in Intensive Care Units (ICUs) can be more efficient through accurate LOS–P predictions since the length-of-stay is an important source of variability in the use of resources in those units. LOS–P forecasting models also enable hospital bed management through intelligent methods, basing admissions on future resource availability [2]. Understanding the variation in LOS allows projections that result in organizational and financial benefits [28]. Zernikow et al. [35] cite that in a low-budget scenario, shift staff scheduling is key for resource planning and may be optimized through the precise modeling of LOS–P.

Jeremic and Tan [41] argue that the development of forecasting tools assists healthcare professionals in making decisions and, within acceptable error levels, allows medical teams to optimize decisions related to treatments and procedures. Decision support systems based on LOS–P forecasting models are relatively low-maintenance and provide real-time information, which helps in hospital management [27], [42].

Works included in this scoping review present some gaps related to the studied population and departments. Most LOS–P forecasting studies are focused on newborn patients; the analysis of child and adolescent patients constitutes a research opportunity. Studies are predominantly carried out at Neonatal Intensive Care units and hospitals or Psychiatric units. There is an opportunity to develop studies in Emergency departments and Pediatric Intensive Care units. Datasets including all pediatric ages and departments of pediatric hospitals are also needed to compare the performance between general and dedicated forecasting models.

The approach used to generate LOS–P forecasting models was predominantly regression analysis, with few studies using Machine Learning or Artificial Intelligence techniques. The digitization of healthcare organizations allows automated data collection on patients' LOS and their potential predictors, creating an environment suitable to implement machine learning-based forecasting models, constituting a promising research direction.

Few studies sought to generalize their proposed forecasting models through the analysis of multicenter datasets. Levin et al. [27] pointed out the need for research in this direction to obtain forecasting models that are both externally validated and generalizable, which is still a gap to be bridged in the literature.

Finally, studies from our *corpus* rarely report the real applicability of the proposed forecasting models, mostly indicating only qualitative benefits. Hintz et al. [38] suggested that future research should focus on applying models in hospitals to improve short and long-term results, shifting the focus from the creation and validation of predictive models. Leon et al. [3] encouraged the use of forecasting models in continuous quality improvement activities. The development of studies reporting the use of LOS–P forecasting models as tools for resource optimization, through the qualitative and quantitative reporting of benefits, remains a promising research direction.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

Funding

The author(s) received no specific funding for this work.

Authors' contributions

NM, FF and MR contributed to the conception and design of the study. NM, FF, and MR contributed to the analysis and interpretation of the extracted data. NM and FF were major contributors in writing the manuscript. GT critically reviewed the manuscript for relevant academic content. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Kampstra, N.A., Zipfel, N., van der Nat, P.B. et al. Health outcomes measurement and organizational readiness support quality improvement: a systematic review. *BMC Health Services Research*, 2018, 18, 1005. DOI: 10.1186/s12913-018-3828-9
2. Ma, F., Yu, L., Ye, L., Yao, D. D., & Zhuang, W. Length-of-Stay Prediction for Pediatric Patients with Respiratory Diseases Using Decision Tree Methods. *IEEE Journal of Biomedical and Health Informatics*, 2020, 24(9), 2651-2662. DOI: 10.1109/jbhi.2020.2973285.
3. Leon, S. C., Snowden, J., Bryant, F. B., & Lyons, J. S. The hospital as predictor of children's and adolescents' length of stay. *Journal of the American Academy of Child & Adolescent Psychiatry*, 2006, 45(3), 322-328. DOI: 10.1097/01.chi.0000194565.78536.bb.
4. Simon, T. D., Berry, J., Feudtner, C., Stone, B. L., Sheng, X., Bratton, S. L., ... & Srivastava, R. Children with complex chronic conditions in inpatient hospital settings in the United States. *Pediatrics*, 2010, 126(4), 647-655. DOI: 10.1542/peds.2009-3266.
5. Lemkin, D.L., Stryckman, B., Klein, J.E., Custer, J.W., Maranda, L., Wood, K.E., Paulson, C., Dezman, Z.D.W. Integrating a safety smart list into the electronic health record decreases intensive care unit length of stay and cost. *Journal of Critical Care*, 2019, 57, 246-252. DOI: 10.1016/j.jcrc.2019.09.016.
6. Lisk, R., Uddin, M., Parbhoo, A., Yeong, K., Fluck, D., Sharma, P., ... & Han, T. S. Predictive model of length of stay in hospital among older patients. *Aging Clinical and Experimental Research*, 2019, 31(7), 993-999. DOI: 10.1007/s40520-018-1033-7.
7. Chen, C. H., Hsieh, J. G., Cheng, S. L., Lin, Y. L., Lin, P. H., & Jeng, J. H. Early short-term prediction of emergency department length of stay using natural language processing for low-acuity outpatients. *The American Journal of Emergency Medicine*, 2020, 38(11), 2368-2373. DOI: 10.1016/j.ajem.2020.03.019.
8. Almashrafi, A., Elmontsri, M., & Aylin, P. Systematic review of factors influencing length of stay in ICU after adult cardiac surgery. *BMC Health Services Research*, 2016, 16(1), 318. DOI: 10.1186/s12913-016-1591-3.
9. Peres, I. T., Hamacher, S., Oliveira, F. L. C., Thomé, A. M. T., & Bozza, F. A. What factors predict length of stay in the intensive care unit? Systematic review and meta-analysis. *Journal of Critical Care*, 2020, 60, 183-194. DOI: 10.1016/j.jcrc.2020.08.003.
10. Atashi, A., Verburg, I. W., Karim, H., Miri, M., Abu-Hanna, A., de Jonge, E., ... & Eslami, S. Models to predict length of stay in the Intensive Care Unit after coronary artery bypass grafting: a systematic review. *The Journal of Cardiovascular Surgery*, 2018, 59(3), 471-482. DOI: 10.23736/S0021-9509.18.09847-6.
11. Hussain, A., & Dunn, K. W. Predicting length of stay in thermal burns: a systematic review of prognostic factors. *Burns*, 2013, 39(7), 1331-1340. DOI: 10.1016/j.burns.2013.04.026.
12. Lu, M., Sajobi, T., Lucyk, K., Lorenzetti, D., & Quan, H. Systematic review of risk adjustment models of hospital length of stay (LOS). *Medical Care*, 2015, 53(4), 355-365. DOI: 10.1097/mlr.0000000000000317.
13. Verburg, I. W. M., Atashi, A., Eslami, S., Holman, R., Abu-Hanna, A., de Jonge, E., ... & de Keizer, N. F. Which models can I use to predict adult ICU length of stay? A systematic review. *Critical Care Medicine*, 2017, 45(2), e222-e231. DOI: 10.1097/ccm.0000000000002054.
14. Seaton, S. E., Barker, L., Jenkins, D., Draper, E. S., Abrams, K. R., & Manktelow, B. N. What factors predict length of stay in a neonatal unit: a systematic review. *BMJ Open*, 2016, 6(10), e010466. DOI: 10.1136/bmjopen-2015-010466.
15. Arksey, H., & O'Malley, L. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 2005, 8(1), 19-32. DOI: 10.1080/1364557032000119616.
16. Tortorella, G. L., Fogliatto, F. S., Mac Cawley Vergara, A., Vassolo, R., & Sawhney, R. Healthcare 4.0: trends, challenges and research directions. *Production Planning & Control*, 2020, 31(15), 1245-1260. DOI: 10.1080/09537287.2019.1702226
17. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine*, 2009, 151(4), 264-269. DOI: 10.7326/0003-4819-151-4-200908180-00135.
18. Rendina, M. C. The effect of telemedicine on neonatal intensive care unit length of stay in very low birthweight infants. In *Proceedings of the AMIA Symposium* (p. 111). American Medical Informatics Association, 1998.
19. Walczak, S., & Scorpio, R. J. Predicting pediatric length of stay and acuity of care in the first ten minutes with artificial neural networks. *Pediatric Critical Care Medicine*, 2000, 1(1), 42-47. DOI: 10.1097/00130478-200007000-00008.
20. Höger, C., Zieger, H., Presting, G., Witte-Lakemann, G., Specht, F., & Rothenberger, A. Predictors of length of stay in inpatient child and adolescent psychiatry: failure to validate an evidence-based model. *European Child & Adolescent Psychiatry*, 2002, 11(6), 281-288. DOI: 10.1007/s00787-002-0290-2.
21. Anderson, J. B., Beekman III, R. H., Border, W. L., Kalkwarf, H. J., Khoury, P. R., Uzark, K., ... & Marino, B. S. Lower weight-for-age z score adversely affects hospital length of stay after the bidirectional Glenn procedure in 100 infants with a single ventricle. *The Journal of Thoracic and Cardiovascular Surgery*,

- 2009, 138(2), 397-404. DOI: 10.1016/j.jtcvs.2009.02.033.
22. Nagarsheth, K. H., Gandhi, S. S., Heidel, R. E., Kurek, S. J., & Angel, C. A mathematical model to predict length of stay in pediatric ATV accident victims. *Journal of Surgical Research*, 2011, 171(1), 28-30. DOI: 10.1016/j.jss.2011.03.063.
23. Pepler, F. T., Uys, D. W., & Nel, D. G. Predicting mortality and length-of-stay for neonatal admissions to private hospital neonatal intensive care units: a Southern African retrospective study. *African Health Sciences*, 2012, 12(2), 166-173. DOI: 10.4314/ahs.v12i2.14.
24. Stewart, S. L., Kam, C., & Baiden, P. Predicting length of stay and readmission for psychiatric inpatient youth admitted to adult mental health beds in Ontario, Canada. *Child and Adolescent Mental Health*, 2014, 19(2), 115-121. DOI: 10.1111/camh.12022.
25. Kavanaugh, B., Studeny, J., Cancilliere, M. K., & Holler, K. A. Neurocognitive predictors of length of stay within a children's psychiatric inpatient program. *Child Neuropsychology*, 2020, 26(1), 129-136. DOI: 10.1080/09297049.2019.1617843.
26. Khoshnood, B., Lee, K. S., Corpuz, M., Koetting, M., Hsieh, H. L., & Kim, B. I. Models for determining cost of care and length of stay in neonatal intensive care units. *International Journal of Technology Assessment in Health Care*, 1996, 12(1), 62-71. DOI: 10.1017/s0266462300009399.
27. Levin, S. R., Harley, E. T., Fackler, J. C., Lehmann, C. U., Custer, J. W., France, D., & Zeger, S. L. Real-time forecasting of pediatric intensive care unit length of stay using computerized provider orders. *Critical Care Medicine*, 2012, 40(11), 3058-3064. DOI: 10.1097/ccm.0b013e31825bc399.
28. Bender, G. J., Koestler, D., Ombao, H., McCourt, M., Alskinis, B., Rubin, L. P., & Padbury, J. F. Neonatal intensive care unit: predictive models for length of stay. *Journal of Perinatology*, 2013, 33(2), 147-153. DOI: 10.1038/jp.2012.62.
29. Pearlman, S. A., Stachecki, S., Aussprung, H. L., & Raval, N. Predicting length of hospitalization of sick neonates from their initial status. *Clinical Pediatrics*, 1992, 31(7), 391-393. DOI: 10.1177/000992289203100702.
30. Gold, J., Shera, D., & Clarkson JR, B. Private psychiatric hospitalization of children: predictors of length of stay. *Journal of the American Academy of Child & Adolescent Psychiatry*, 1993, 32(1), 135-143. DOI: 10.1097/00004583-199301000-00020.
31. Pastura, P. S. V. C., Land, M. G., & Santoro-Lopes, G. Predictive model for the length of hospital stay of children with hematologic malignancies, neutropenia, and presumed infection. *Journal of Pediatric Hematology/Oncology*, 2004, 26(12), 813-816.
32. Paul, M., Partridge, J., Barrett-Reis, B., Ahmad, K. A., Machiraju, P., Jayapalan, H., & Schanler, R. J. Metabolic Acidosis in Preterm Infants is Associated with a Longer Length of Stay in the Neonatal Intensive Care Unit. *PharmacoEconomics-Open*, 2020, 4(3), 541-547. DOI: 10.1007/s41669-020-00194-y.
33. Parkman, S. E., & Woods, S. L. Infants who have undergone cardiac surgery: what can we learn about lengths of stay in the hospital and presence of complications?. *Journal of Pediatric Nursing*, 2005, 20(6), 430-440. DOI: 10.1016/j.pedn.2005.03.013.
34. Balan, S., Gawade, T., & Tasgaonkar, A. A Machine Learning Approach for Prediction of Length of Stay for the Kid's Inpatient Database. In 2019 IEEE International Conference on Big Data (Big Data), 2019, p. 5980-5982. DOI: 10.1109/EMBC44109.2020.9175889.
35. Zernikow, B., Holtmannspötter, K., Michel, E., Hornschuh, F., Groote, K., & Hennecke, K. H. Predicting length-of-stay in preterm neonates. *European Journal of Pediatrics*, 1999, 158(1), 59-62. DOI: 10.1007/s004310051010.
36. Bannwart, D., Rebello, C. M., dos SR Sadeck, L., Pontes, M. D., Ramos, J. L. A., & Leone, C. R. Prediction of length of hospital stay in neonatal units for very low birth weight infants. *Journal of Perinatology*, 1999, 19(2), 92-96. DOI: 10.1038/sj.jp.7200134.
37. Browning, D. L. Psychiatric ward behavior and length of stay in adolescent and young adult inpatients: A developmental approach to prediction. *Journal of Consulting and Clinical Psychology*, 1986, 54(2), 227-230. DOI: 10.1037//0022-006x.54.2.227.
38. Hintz, S. R., Bann, C. M., Ambalavanan, N., Cotten, C. M., Das, A., & Higgins, R. D. Predicting time to hospital discharge for extremely preterm infants. *Pediatrics*, 2010, 125(1), e146-e154. DOI: 10.1542/peds.2009-0810.
39. Lee, H. C., Bennett, M. V., Schulman, J., Gould, J. B., & Profit, J. Estimating length of stay by patient type in the neonatal intensive care unit. *American Journal of Perinatology*, 2016, 33(08), 751-757. DOI: 10.1055/s-0036-1572433.
40. Marshall, G., Luque, M. J., Gonzalez, A., Musante, G., & Tapia, J. L. Center variability in risk of adjusted length of stay for very low birth weight infants in the Neocosur South American Network. *Jornal de Pediatria*, 2012, 88(6), 524-530. DOI: 10.2223/jped.2234.
41. Jeremic, A., & Tan, K. (2008). Predicting the length of stay for neonates using heart-rate Markov models. In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008, p. 2912-2915. DOI: 10.1109/IEMBS.2008.4649812.
42. Walsh, P., Cunningham, P., Rothenberg, S. J., O'Doherty, S., Hoey, H., & Healy, R. An artificial neural network ensemble to predict disposition and length of stay in children presenting with bronchiolitis. *European Journal of Emergency Medicine*, 2004, 11(5), 259-264. DOI: 10.1097/00063110-200410000-00004.
43. Lee, A. H., Gracey, M., Wang, K., & Yau, K. K. A robustified modeling approach to analyze pediatric length of stay. *Annals of Epidemiology*, 2005, 15(9), 673-677. DOI: 10.1016/j.annepidem.2004.10.001.
44. Baniyasi, T., Kahnouji, K., Davaridolatbadi, N. et al. Factors affecting length of stay in Children Hospital in Southern Iran. *BMC Health Services Research*, 2019, 19, 949. DOI: 10.1186/s12913-019-4799-1

Figures

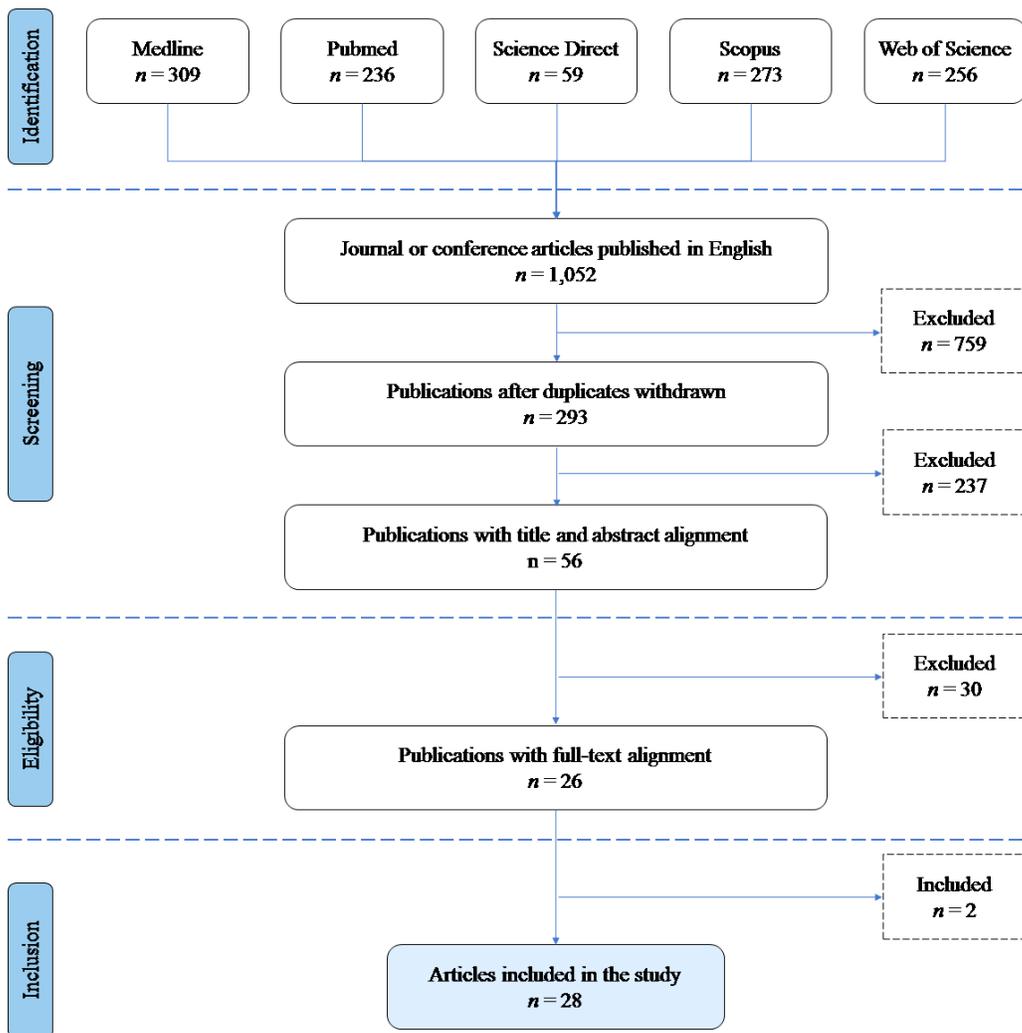


Figure 1

Results of PRISMA stages

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [PRISMAchecklist.doc](#)
- [PRISMAflowdiagram.doc](#)