

Types of statistical analysis in prognostic studies

Mario E. Rendón-Macías*, Ana S. Castillo-Ivón, and Lorena Orozco-Díaz

Faculty of Health Sciences, School of Medicine, Universidad Panamericana, Mexico City, Mexico

Abstract

Prognostic studies may have a descriptive exploratory objective on an outcome or a comparative objective in the search for factors associated with it. A second objective is explanatory to determine the effect of a particular prognostic factor adjusted for its confounders, with or without the intention of establishing causality. The third objective is the construction of a predictive prognostic scale. For each of these objectives, there are recommended statistical methods for clarification and validity. In this article, the methods and application examples are presented. The proper selection of analytical methods allows for clear and valid communication of the results of a prognostic study.

Keywords: Statistical analysis. Prognosis. Validity.

Tipos de análisis estadísticos en los estudios de pronóstico

Resumen

Los estudios pronósticos pueden tener un objetivo exploratorio descriptivo sobre un desenlace o comparativo en la búsqueda de factores asociados al mismo. Un segundo objetivo es explicativo para determinar el impacto de un factor pronóstico en particular ajustado a sus confusores con o sin la intención de establecer rutas causales. El tercero es la construcción de una escala pronóstica predictiva. Para cada uno de estos objetivos existen métodos estadísticos recomendados para su clarificación y validez, los cuales fueron revisados en una publicación previa. En este artículo presentamos los métodos y ejemplos de aplicación. La adecuada selección de los métodos analíticos permite una comunicación clara y válida de los resultados de un estudio pronóstico.

Palabras clave: Análisis estadístico. Pronóstico. Validez.

***Correspondence:**

Mario E. Rendón-Macías

E-mail: mrendon@up.edu.mx

1665-1146/© 2024 Hospital Infantil de México Federico Gómez. Published by Permanyer. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Date of reception: 01-05-2024

Date of acceptance: 02-08-2024

DOI: 10.24875/BMHIM.24000060

Available online: 21-11-2024

Bol Med Hosp Infant Mex. 2024;81(6):319-327

www.bmhim.com

Introduction

Prognostic studies analyze the potential consequences of suffering from a disease and can be classified according to three general objectives: exploratory, explanatory, or predictive¹. Exploratory studies aim to establish the probability of occurrence of relevant outcome(s) in the studied patients. For explanatory studies, the intention is to validate the independent effect of a particular factor of interest on the outcome(s), adjusted for known confounding factors. Studies with the third objective aim to construct a prognostic prediction scale, as precise as possible, based on patient data¹⁻⁴. Each purpose requires a distinct methodology to obtain valid and useful data for reliable statistical analysis¹. During the reading, review, or execution of a prognostic study, it is common to find readers with doubts about the recommended statistical procedures according to the objectives mentioned above. This review analyzes the recommended statistical strategies for data analysis for the different prognostic objectives.

Prognostic studies with exploratory purpose

In this type of study, the analysis can be conducted in a descriptive or comparative manner (Fig. 1). Descriptive analysis aims to report the frequency and proportion (or percentage) of patients who developed the outcome under study (e.g., mortality or a sequela). For this information, the patient follow-up method must be considered. If all patients had the same follow-up time (e.g., 24 h), it is only necessary to present the cumulative incidence of the outcome(s). For example, in a study on the prognosis of intubation in patients with severe asthma attacks, one can report that 20% of individuals admitted to the emergency room ended up receiving ventilation assistance within 24 h after admission. In this analysis, it is feasible to report on several outcomes (e.g., fatality or ventilator-associated pneumonia, among others). If, in addition, one wishes to consider the rate at which the outcome occurs, it can be reported as an incidence rate (events per person-time). For the first option, actuarial tables are used; for the second, survival tables and Kaplan–Meier curves are adequate (Table 1 and Fig. 2)⁵.

If one also wishes to establish whether any factor present at the beginning of the clinical course follow-up (initial cohort) could explain the different outcome(s), the first approach is to observe the proportion

Table 1. Actuarial and survival table of the need for orotracheal intubation in patients with asthmatic crisis (fictitious data of n=135 persons)

Actuarial analysis			
Timing	Intubated (n)	% Events in the period	% Cumulative survival
Admission	0	0	100
6 h	5	3.7 (5/135)	96.3
12 h	6	4.7 (6/130) 17.7	91.8
18 h	22	(22/124)	75.5
24 h	12	11.7 (12/102)	66.7
30 h	6	6.7 (6/90)	62.2
36 h	2	2.4 (2/84)	60.7
42 h	3	3.6 (3/82)	58.5
48 h	1	1.2 (1/81)	57.8
Person-time survival analysis			
Timing	Intubated (n)	% Events in the period	% Cumulative survival
Admission	0	0	100
2 h	1	0.7 (1/135)	93
3 h	5	3.7 (5/134)	89.6
6 h	(1 lost)	-	89.6
9 h	3	2.3 (3/133)	87.5
19 h	5 (1 lost)	3.1 (4/130)	84.8
22 h	10	8 (10/125)	78
25 h	2	1.7 (2/115)	76.7
30 h	(1 lost)	-	76.7
39 h	1	0.8 (1/114)	76.1
41 h	15	13.3 (15/113)	65.9
48 h	3	2.7 (3/110)	64.2

The percentage of events is the number of events presented in the period among patients not yet intubated. The probability of not being intubated is the product of the probability of remaining without intubation in the previous period and the probability of remaining without being intubated in the analysis period. In the actuarial table, the periods are fixed; in the survival table, it is recorded when at least one event occurs or if the follow-up of at least one patient is lost.

of subjects with this factor among those who did or did not present the studied outcome(s). In the case that the follow-up time is the same for all patients, it is sufficient to compare their cumulative incidence rates using a test of difference in proportions (for example, the Chi-square test) or the 95% confidence intervals (CI) of the differences in proportions (if the interval includes the value “0”, it is not statistically conclusive)⁶.

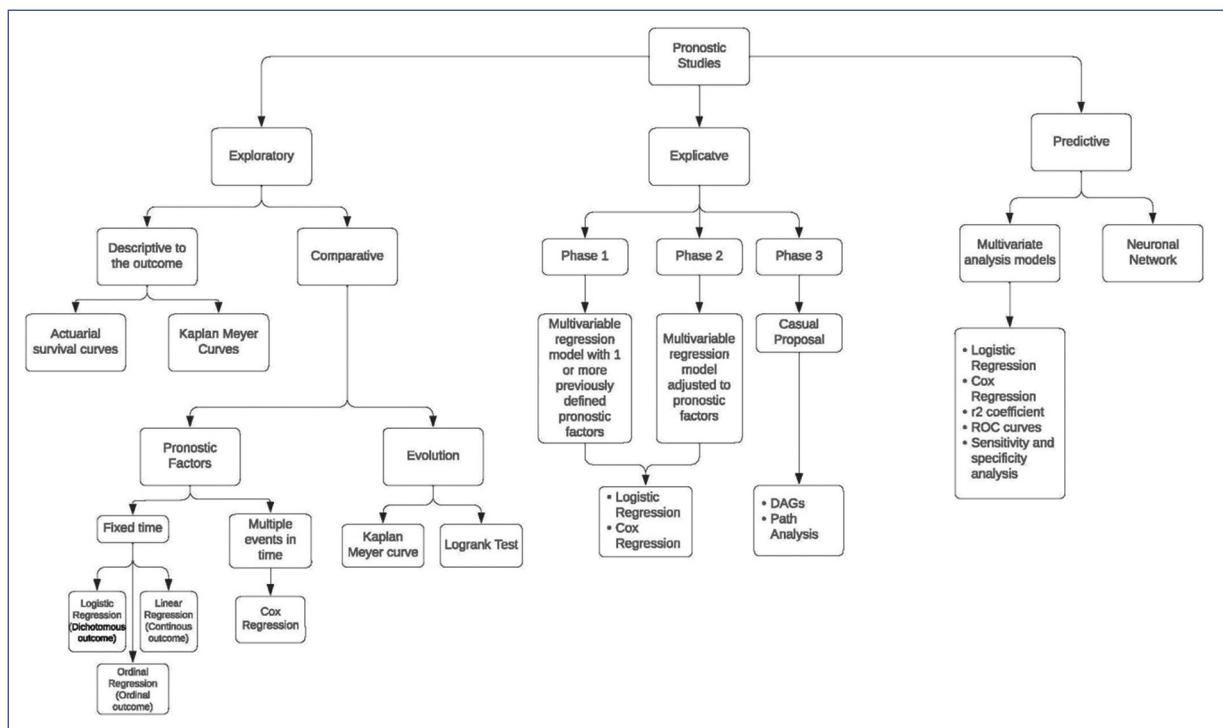


Figure 1. Diagram of recommended statistical methods according to the objective of the prognostic study.

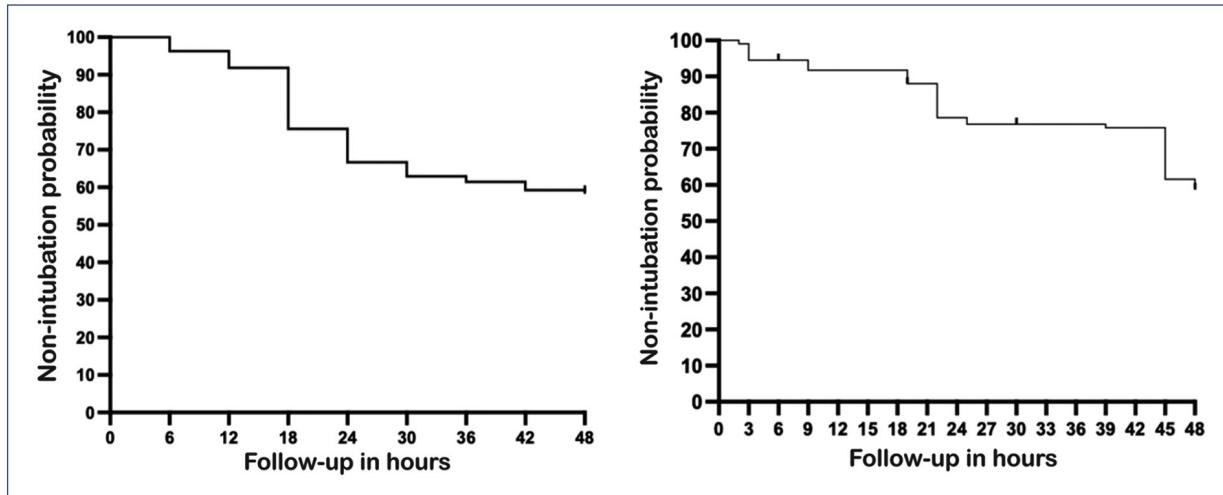


Figure 2. Actuarial curves with fixed analysis times (equal intervals) versus the Kaplan-Meier survival curve where it decreases in the presence of at least one event in the real follow-up time, the vertical mark indicates the censoring of at least one patient (loss to follow-up without presenting the event). The presented data are fictitious and obtained from [table 2](#).

Another approach involves comparing the velocities of outcomes between patients with and without the factor. The test of choice is the “log-rank test”^{5,7}. For example, the median intubation-free survival was 12 h

for patients with atopy, compared to 20 h for those without atopy (mean difference of -8 h, 95% CI from -12 to -6 h, log-rank test $p = 0.001$, data calculated as an example).

A severe problem in bivariate comparisons (groups with and without the prognostic factors to be evaluated) is that, in some cases, statistically significant differences can be found due to multiple possible comparisons. This is due to the increased risk of committing a type I error (bias due to “multiple comparisons”)⁸. These models assume the possibility of knowing how much a factor influences the outcome, considering the partial effect of others (adjustment), that is, how much the factor influences independently of another or others. The correct way to jointly analyze several factors to establish which one(s) are associated with the outcome and review which one(s) are more influential is through multivariable regression models^{9,10}. The choice of model will depend on how the outcome variable was measured and the form of follow-up (at fixed or continuous times), as well as verifying compliance with a series of statistical assumptions necessary to establish its validity (Table 2)^{9,10}. For outcomes with fixed times, the most used regression models are binary logistic (dichotomous outcome: presence or absence of the outcome), multiple linear (quantitative outcome: days of hospitalization), or ordinal (hierarchical qualitative outcome: mild, moderate, and severe damage)¹⁰.

The interpretation is based on the beta coefficients of each model. In logistic and ordinal regression, the exponential of beta or odds ratio (OR) is used. They are from zero to infinity and the null value is “1”, the further away from 1 the greater association. If the CI does not include it, it will be significant at the established level (90, 95, or 99%) (Table 2a). In multiple linear regression, the comparison is made with the values of the standardized beta coefficients, which eliminates the original unit of measurement and allows for comparability of the effect of each factor. In this model, the null hypothesis of no association is the existence of a standardized coefficient with a value of “0”. The further it deviates ($-\infty$ or $+\infty$), the greater the impact it will have on the prognosis. If the X% CI includes the value of “0,” the result will not be significant at the established level, or there will be no association^{11,12} (Table 2b).

In these models, as for the multivariable linear model, the following assumptions are considered: linearity between predictors and outcome, homoscedasticity, normality and independence of residuals, and multicollinearity or high correlation among predictor factors. In logistic regression, it is mainly concerned with avoiding multicollinearity and independence in individual exposure to factors. When linearity does not

exist, scale transformation options or stepwise analyses may be employed to facilitate the analysis, although it is recommended to consult with a statistical expert to avoid losing clinical significance. Multicollinearity is the second major problem in multivariable analysis; to avoid it, we recommend carefully reviewing the factors to be considered, and when there is a high correlation among some of them, consider including in the model only the factor with better measurement, greater validity, stronger association with the outcome, and less loss or absence in its capture.

When the outcome is a proportion adjusted for the time of presentation, the recommended model is Cox regression^{13,14}. This model assumes that the risk(s) are always continuous and proportional (proportional hazards assumption), so the beta coefficient is presented as a hazard ratio (HR). It is also necessary to meet the assumptions of the absence of multicollinearity, linearity in the predictor variables with the logarithm of the outcome rate, and the absence of outliers. The interpretation of an HR is similar to that of an OR, that is, how many times more or less likely the presence of the complication is when exposed to a factor compared to not being exposed to it (Table 2c)¹²⁻¹⁴. In all the above models, researchers should report on statistically significant factors and highlight those with more extreme values concerning the null value. Other multivariable regression models are not mentioned here; interested readers are advised to consult statistical professionals.

Prognostic studies with explanatory purpose

As mentioned earlier, the objective is to validate the impact of a prognostic factor of interest controlled by its possible confounders. It should be remembered that a confounding factor is one known to be causal of the outcome of interest but associated with the prognostic factor under study without being part of the pathophysiological pathway by which the factor under study explains the outcome. In this analysis, it is also recommended to perform a multivariable regression with the same specifications mentioned previously. The main difference is that only the prognostic factor of interest and its confounders should be included in the model, not just any factor. It is important to select confounders adequately because as they increase, it will be necessary to expand the sample size^{11,15,16}.

Table 2. Prognostic factors associated with relapse of urticaria syndrome (fictitious data)

a) Example of logistic regression analysis, risk of relapse 1 month after resolution			
Factors	OR (Exponent of beta)	(95% CI, lower-upper limit)	p-value*
History of allergy	3.5	(2.1 a 4.3)	0.001
Use of antihistamines	0.4	(0.35 a 0.6)	0.02
Female sex	1.2	(0.8 a 1.6)	0.83
Age under 18 years	2.1	(0.3 a 5.2)	0.45
Nutritional status			
Obesity	1.9	(0.7 a 2.2)	0.55
Overweight	1.4	(0.9 a 3.1)	0.67
Adequate weight	reference		

Associated factors are a history of allergy and use of antihistamines, the former with a greater risk impact and the latter with a moderate protective or preventive effect.

*Wald statistical test. Null value = 1.

b) Example of linear regression analysis, risk of lesion persistence, number of days			
Factors	Standardized beta	(95% CI of standardized betas)	p-value*
History of allergy	1.2	(0.9 to 1.4)	0.002
Use of antihistamines	-0.6	(-0.3 to -0.8)	0.031
Female sex	0.03	(-0.8 to 1.8)	0.83
Age under 18 years	0.01	(-0.3 to 0.5)	0.51
BMI	0.03	(-0.02 to 0.02)	0.87

Associated factors are: history of allergy and use of antihistamines, the former with a greater positive association (having the history means more days of persistence), and the latter with a moderate reducing (inverse) effect on the days. *Student's t-test. Null value = 0.

c) Example of Cox regression analysis, continuous risk of relapse			
Factors	Hazard ratio	(95% CI, lower-upper limit)	p-value*
History of allergy	3.3	(2 a 4.2)	0.002
Use of antihistamines	0.39	(0.31 a 0.6)	0.02
Female sex	1.02	(0.75 a 1.7)	0.83
Age under 18 years	2.2	(0.31 a 5.2)	0.46
Nutritional status			
Obesity	1.8	(0.7 a 2.3)	0.57
Overweight	1.3	(0.8 a 3.2)	0.69
Adequate weight	reference		

Associated factors are a history of allergy and use of antihistamines, the former with a greater risk impact, and the latter with a moderate protective or preventive effect. *Wald statistical test. Null value = 1. CI: Confidence intervals; OR: Odds ratio; BMI: Body mass index.

We suggest including the most involved, prevalent, better-measured confounders, with the potential to be modifiable in the future and the easiest to obtain⁴. In the final analysis report, the association estimator between the studied prognostic factor (relative risk, OR, HR, or standardized beta) and the outcome (e.g., relapse rate) should be shown, indicating the confounding factors to which the association was adjusted. It does not make sense to report on the estimators of

the confounders since these were not adjusted for their own confounders and, therefore, they do not have explanatory value. If the factor of interest is removed, the study loses its objective. An example of a report is presented in [table 3](#).

A proposed phase for this purpose is the causal network⁴. In this model, the factor under study and its outcome are not only adjusted for confounders analyzed but also antecedent and modifying factors are

Table 3. Prognostic factors associated with relapse of urticaria syndrome (fictitious data)

a) Example of logistic regression analysis. History of allergy as a prognostic factor for relapse 1 month after resolution			
Factors	OR (exponent of beta)	(95% CI, lower-upper limit)	p-value*
History of allergy	3.5	(2.1 a 4.3)	0.001

Adjusted for use of antihistamines, sex, age, and nutritional status

b) Example of linear regression analysis. Effect of history of allergy as a prognostic factor for the duration of lesion persistence in number of days			
Factors	Standardized beta	(95% CI of standardized betas)	p-value**
History of allergy	1.2	(0.9 a 1.4)	0.002

Adjusted for use of antihistamines, sex, age, and nutritional status.

c) Example of Cox regression analysis. History of allergy as a prognostic factor for continuous risk of relapse			
Factors	Hazard ratio	(95% CI, lower-upper limit)	p-value*
History of allergy	3.3	(2 a 4.2)	0.002

Adjusted for use of antihistamines, sex, age, and nutritional status.

*Wald statistical test, p-value.

**T statistical test, p-value.

CI: Confidence intervals; OR: odds ratio

also added. Directed acyclic graphs models and path analysis have been proposed for its presentation^{3,4}. Given their limited use in clinical medicine, readers are invited to consult specific sources⁴.

Prognostic studies with predictive purpose

These models are created to generate diagnostic and prognostic scales. In general, it is recommended to analyze these models in three phases: construction, internal validation, and external validation. In this review, we will only refer to their internal validation. For validation, two main types of analysis are primarily used: multivariable regression models and neural network models. In the former, modeling with regression analysis again depends on the type of dependent variable. The difference lies in the construction of the model. The objective of the selected model is based on one that is (1) more predictive, (2) parsimonious, (3) simple to apply, and (4) universal^{9,11,16,17}.

To validate a predictive model based on multivariate regression, it is necessary to consider a large sample size, generally ten patients for each factor to be considered. Once the sample is available, the analysis is executed with a statistical computer program. Regardless of the program used, it will request a dependent variable (the outcome of interest) and the introduction of independent variables or covariates. The objective of the analysis is to find an equation that allows obtaining (predicted) values as close as possible to those observed in patients

(real). If this approximation is excellent, it will generally be excellent for patients with similar conditions who did not participate in the equation validation study (external validity). The prediction can be in terms of the probability of an outcome, time to an event, time to an outcome, and level of severity, among others. The method of selecting the most predictive variables of the outcome is based on the amount of variation explained by the equation. The most used estimator to address this situation is the coefficient of determination or R^2 (pseudo R^2 for logistic regression). The R^2 coefficient ranges from zero, which predicts nothing, to one, which implies a perfect prediction. To find the variables that will generate the most predictive equation, computers use three methods: forward, backward, or stepwise (Fig. 3 and Table 4). In the first method, all proposed variables are reviewed, and the most significant in its association with the outcome is selected (for example, the smallest "p" value). Next, the second most significant is sought, and if a significant change in R^2 is found, a third most associated factor is added. This process is repeated until no significant improvement in R^2 is observed, indicating a lack of predictive gain with more factors (Table 4a). The second method performs the procedure in the opposite way. It begins by introducing all the factors considered and calculating R^2 . Then, it eliminates non-significant (associated) factors one by one and reviews the R^2 coefficient, which does not reduce the prediction. When removing a factor causes R^2 to decrease, the program stops subtracting factors, and the remaining ones are those that

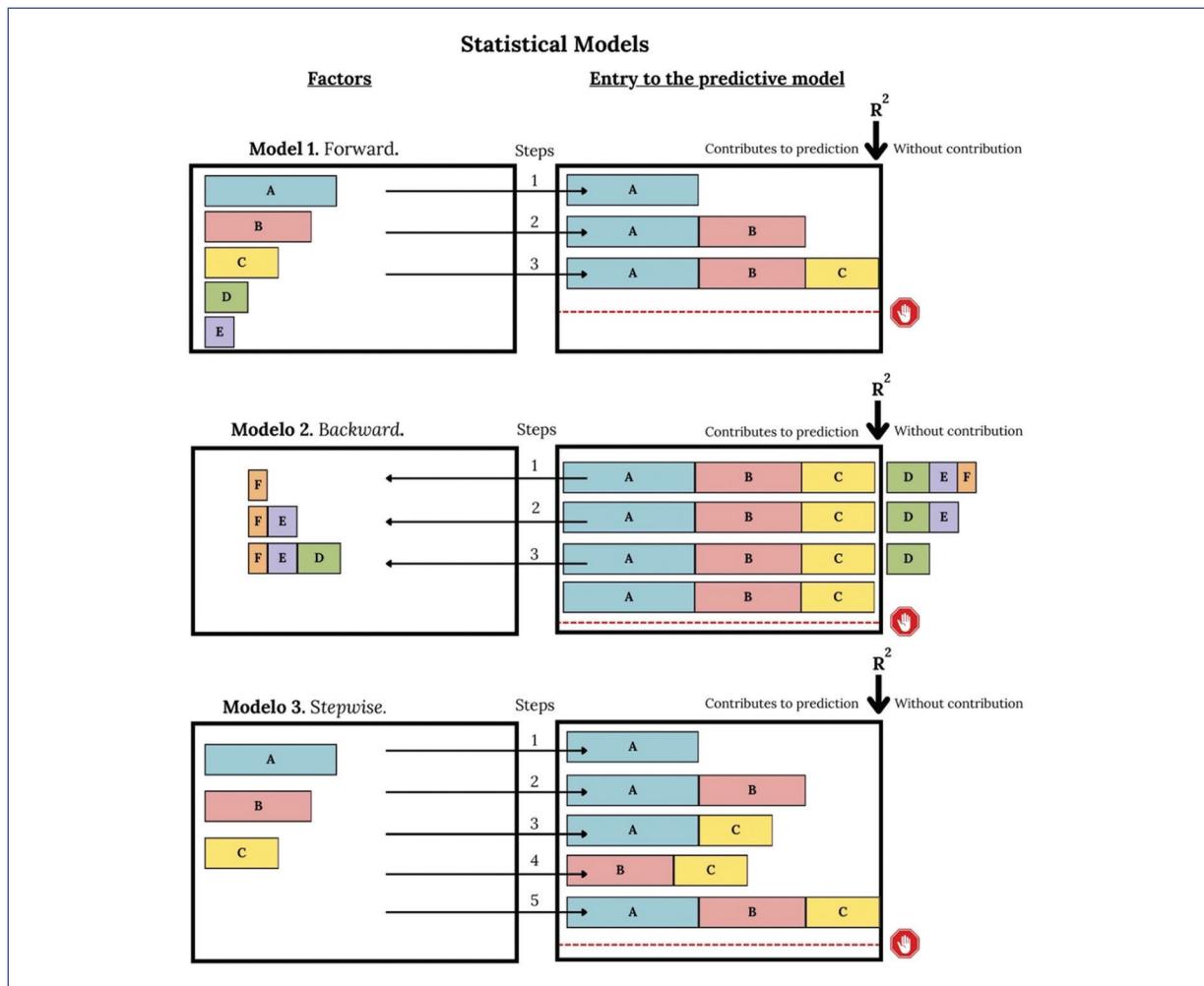


Figure 3. Statistical modeling options to obtain the most precise prediction model. The squares represent prognostic variables and their size indicates the level of association with the prognostic variable. The R^2 value informs about the maximum prediction range.

provide the greatest prediction (Table 4b). The third method (stepwise) is the most recommended. The selection is based on conducting trials of incorporating and removing factors in search of the combination with the highest coefficient of determination, that is, the most predictive (Table 4c).

In prognostic scales where the outcome is quantitative (for example, days of hospital stay or years of survival), it is only necessary to establish the best prediction equation. However, if the outcome variable is qualitative (for example, cure), the programs determine the probability of the event as present if the constructed equation gives a score of 0.5 or more (50% or more). It is possible to improve the interpretation of diagnostic and prognostic validity by estimating its highest sensitivity and specificity by constructing a receiver operating characteristic curve and its area under the curve (Fig. 4). It is also

feasible to determine the degree of discrimination of the prediction equation through specific analyses¹⁸. Alongside the validation of the most predictive model, it is necessary to consider other criteria. Parsimony refers to the model that has fewer included factors. In general, a model with more factors considered allows for better prediction. However, its use can become complicated if more than ten are included, given the difficulty in memorizing them or the lack of availability of information on some occasions. If a model with fewer factors does not significantly reduce the prediction by more than 10%, it will be more recommendable. Simplicity refers to having factors that can be determined or measured with unsophisticated methods in terms of cost, time, and execution. Universality implies that the factors can be determined or measured with unsophisticated

Table 4. Predictive models of allergic dermatitis at 1 year of life in neonates with intolerance to breast milk according to model types (fictitious data n = 416)

a) Example of logistic regression analysis. Forward model			
Factors	beta	p-value*	Pseudo-R ² **
Model 1			
Birth weight (g)	0.004	< 0.001	
Constant	10.8	< 0.001	0.57
Model 2			
Family atopy	2.01	0.002	
Birth weight (g)	-0.004	< 0.001	
Constant	10.8	< 0.001	0.59
Model 3			
Vaccination reaction	1.67	0.016	
Family atopy	2.1	0.001	
Birth weight (g)	-0.004	< 0.001	
Constant	10.7	< 0.001	0.61
b) Example of logistic regression analysis. Backward model			
Factors	Beta	p-value*	Pseudo-R ² **
Model 1			
Birth weight (g)	-0.004	< 0.001	
Family atopy	1.9	0.004	
Vaccination reaction	1.56	0.027	
Iron intake	0.024	0.51	
Calcium intake	0.01	0.53	
Constant	10.8	< 0.001	0.609
Model 2			
Birth weight (g)	1.2	< 0.001	
Family atopy	2.01	0.003	
Vaccination reaction	1.57	0.025	
Iron intake	0.25	0.61	
Constant	9.8	< 0.001	0.608
Model 3			
Birth weight (g)	-0.004	< 0.001	
Family atopy	2.11	0.001	
Vaccination reaction	1.7	0.16	
Constant	10.8	< 0.001	0.83
c) Example of logistic regression analysis. Stepwise model			
Factors	beta	p-value*	Pseudo-R ² **
Final model (3 steps)			
Birth weight (g)	-0.004	< 0.001	
Family atopy	2.12	0.001	
Vaccination reaction	1.7	0.016	
Constant	10.8	< 0.001	0.607

Prognostic factors considered were birth weight, family atopy, vaccination reaction, iron intake, and calcium intake.

*Wald statistical test, *p-value. **Pseudo-R² of Nagelkerke.

methods in terms of cost, time, and execution, which will allow their application in different settings.

Finally, neural network models are based on learning algorithms to obtain the best predictions. Computer systems functions such as the human mind, receiving information continuously, and determining the pathways that

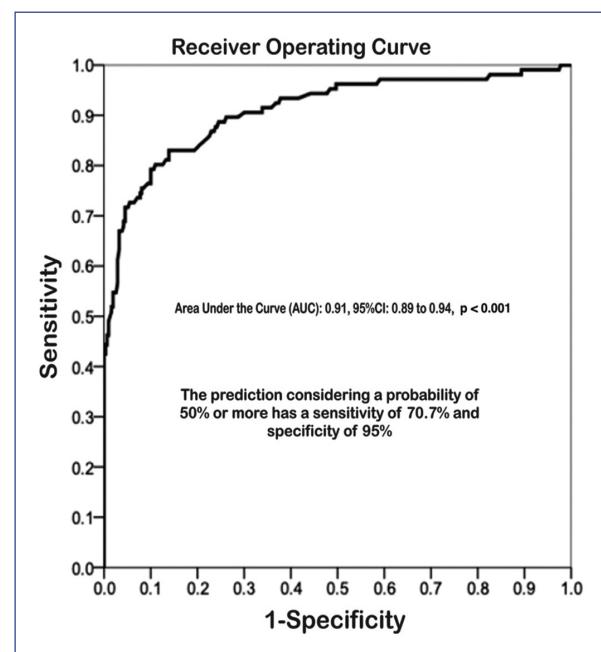


Figure 4. Receiver operating characteristic curve of predictive validity of equation obtained in the analysis of table 4c.

facilitate the approach to a result or “output” with layers or connection capacity. These models are gaining much acceptability due to their high level of prediction¹⁹⁻²⁰. However, they work as “black boxes,” where the connections and functions related to this prediction are unknown, and they are not exempt from methodological biases²¹. To develop them, it is necessary to have the support of specialists in the field, and their validation never ends, given that the more information, the better the prediction. On the other hand, they are not exempt from the criteria mentioned above for simplicity and availability of information.

Conclusion

The recommended statistical analyses in prognostic studies vary according to their objective. These analyses can be merely descriptive, comparative, exploratory, explanatory, or prediction models. The most used methods are multivariable regressions, which are executed and reported according to the objective of the prognostic study. We always recommend seeking advice from a professional in the corresponding area and a statistician to achieve the proposed objective and communicate the results more efficiently.

Funding

The authors declare that they have not received funding.

Conflicts of interest

The authors declare no conflicts of interest.

Ethical considerations

Protection of human and animal subjects. The authors declare that no experiments were performed on humans or animals for this study.

Confidentiality of data. The authors declare that they have followed the protocols of their work center on the publication of patient data.

Right to privacy and informed consent. The authors have obtained the written informed consent of the patients or subjects mentioned in the article. The corresponding author has this document.

References

- Rendón-Macías ME, Castillo-Ivón AS. Methodology for the elaboration of prognosis studies. *Rev Alerg Mex.* 2022;69:48-55.
- Hayden JA, Côté P, Steenstra IA, Bombardier C, QUIPS-LBP Working Group. Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. *J Clin Epidemiol.* 2008;61:552-60.
- Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ.* 2009;338:b375.
- Kent P, Cacchieri C, Boyle E, Cassidy D, Kongsted A. A conceptual framework for prognostic research. *BMC Med Res Methodol.* 2020;20:172.
- Idrayan A, Bansal AK. The methods of survival analysis for clinicians. *Indian Pediatr.* 2010;47:743-8.
- Martínez-Ezquerro JD, Riojas-Garza A, Rendón-Macías ME. Significancia clínica sobre significancia estadística. Cómo interpretar los intervalos de confianza a 95 [Clinical significance vs statistical significance. How to interpret the confidence interval at 95]. *Rev Alerg Mex.* 2017;64:477-86.
- Watanabe H. Applications of statistics to medical science, IV survival analysis. *J Nippon Med Sch.* 2012;79:176-81.
- McHugh ML. Multiple comparison analysis testing in ANOVA. *Biochem Med (Zagreb).* 2011;21:203-9.
- Katz M, editor. Studies of diagnostic and prognostic tests (predictive studies). In: *Study Design and Statistical Analysis: A Practical Guide for Clinicians.* United Kingdom: Cambridge University Press; 2024. p. 141-54.
- Craddock M, Crockett C, McWilliam A, Price G, Sperrin M, van der Veer SN, et al. Evaluation of prognostic and predictive models in the oncology clinic. *Clin Oncol (R Coll Radiol).* 2022;34:102-13.
- Steyerberg EW, Vickers AJ, Cook NR, Gerts T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21:128-38.
- Rendón-Macías ME, Zarco-Villavicencio IS, Villasis-Keever MÁ. *Métodos estadísticos para el análisis del tamaño del efecto* [Statistical methods for effect size analysis]. *Rev Alerg Mex.* 2021;68:128-36. [Spanish].
- Crichton N. Cox proportional hazards model. *J Clin Nurs.* 2002;11:723.
- Pérez-Rodríguez M, Rivas-Ruiz R, Palacios-Cruz L, Talavera JO. *Investigación Clínica XXII. Del juicio clínico al modelo de riesgos proporcionales de Cox* [Clinical research XXII. From clinical judgment to Cox proportional hazards model]. *Rev Med Inst Mex Seguro Soc.* 2014;52:430-5.
- Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. *BMJ.* 2009;339:b4184.
- Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162:W1-73.
- Han K, Song K, Wook-Choi B. How to develop, validate, and compare clinical prediction models involving radiological parameters: study design and statistical methods. *Korean J Radiol.* 2016;17:339-50.
- Tjur T. Coefficients of determination in logistic regression models--A new proposal: the coefficient of discrimination. *Am Stat.* 2009;63:366-372.
- Deo RC. Machine learning in medicine. *Circulation.* 2015;132:1920-30.
- Wolk DM, Lanyado A, Tice AM, Shermohammed M, Kinai Y, Goren A, et al. Prediction of influenza complications: development and validation of a machine learning prediction model to improve and expand the identification of vaccine-hesitant patients at risk of severe influenza complications. *J Clin Med.* 2022;11:4342.
- Andaur-Navarro CL, Damen JA, Takada T, Nijman SW, Dhiman P, Collins GS, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ.* 2021;375:n2281.