Pesquisa Operacional (2023) 43: e275795 p.1-17
doi: 10.1590/0101-7438.2023.043.00275795

Printed version ISSN 0101-7438 / Online version ISSN 1678-5142
www.scielo.br/pope
ARTICLES

PREDICTION AND FACTORS RELATED TO NO-SHOW: A STUDY IN A CARDIOLOGY AND NEUROLOGY CLINIC

Igor Tona Peres[1*], Leila Figueiredo Dantas[2], Davi Melhoranse Gouveia[3] and Fabio Schneider[4]

Received June 20, 2023 / Accepted September 28, 2023

ABSTRACT. Background: The no-show of patients to their scheduled appointments has caused a large increase in healthcare costs, worsening service quality and clinical efficiency. **Objectives:** This case study aims to identify the factors associated with patient no-shows in cardiology and neurology clinics, and develop a prediction model to estimate the no-show probability. **Methods:** We developed a retrospective analysis of 32,573 appointments from January 2019 to June 2022 in a Rio de Janeiro clinic. Logistic regressions were performed to analyze and model the influence of patient and appointment variables on no-show rates. **Results:** The factors most related to no-shows were the patient's sex, age, medical specialty, month of the year, and type of insurance. Female patients have an increase of approximately 17% chance of no-shows compared to males. The age group with the highest no-show rates is between 21 and 30. Clinic consultations have higher no-shows when compared to medical procedures. Appointments in December tend to have higher non-attendance than in January, and patients with insurance from the five major companies presents greater no-show than those with smaller insurance. The prediction model presented the following performance indicators: AUC = 0.65, Sensitivity = 0.64, Specificity = 0.58, PPV = 0.11, and NPV = 0.95. **Conclusions:** This work contributes to understanding the factors related to non-attendance, assisting optimized management of appointment schedules.

Keywords: no-show, appointments, clinical services, logistic regression, prediction model.

---

*Corresponding author

[1]Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro, Rua Marquês de São Vicente, 225, Rio de Janeiro, RJ, Brazil, 22451-900 – E-mail: igor.peres@puc-rio.br – http://orcid.org/0000-0001-9863-4690

[2]Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro, Rua Marquês de São Vicente, 225, Rio de Janeiro, RJ, Brazil, 22451-900 – E-mail: leilafdantas@tecgraf.puc-rio.br – http://orcid.org/0000-0001-8910-0961

[3]Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro, Rua Marquês de São Vicente, 225, Rio de Janeiro, RJ, Brazil, 22451-900 – E-mail: davimelhoranse@gmail.com – http://orcid.org/0009-0005-2278-6600

[4]Brazilian Society of Cardiology, Av. Marechal Câmara, 160, Rio de Janeiro, RJ, Brazil, 20021-380 – E-mail: fabioschneider9@yahoo.com.br – http://orcid.org/0009-0009-2620-7898

1 INTRODUCTION

A no-show occurs when a patient is informed to prepare for an event but does not attend without prior cancellation (Gupta and Wang, 2012; Tuso et al., 1999). It is one of the main challenges for scheduling in health clinics and has been an obstacle for both provider and patient. To the provider, no-shows decrease productivity and efficiency, reducing revenue and limiting the effective capacity of the clinic. They are tied to the physician's idleness and the need to relocate instead of scheduling a potential new patient (Goffman et al., 2017).

Non-attendance can negatively affect patients, leading to long scheduling lead times and perception of the overall decrease in service quality (Cayirli and Veral, 2003; Parikh et al., 2010). There is also an increase in health complications since the lack of the previously performed diagnosis and the effective medical non-follow-up are results collected by the non-attendance at the clinic and corroborate the worsening of certain diseases (Coelho et al., 2005; LaGanga and Lawrence, 2007; Topuz et al., 2018).

To improve the effectiveness of the clinical services area, some literature studies were conducted trying to understand the no-shows from different approaches. A first strategy would be to evaluate other scheduling policies, using overbooking practices, in which more patients are scheduled to compensate for the no-show (Bhattacharjee and Ray, 2016; Kopach et al., 2007; Peres et al., 2019; Samorani and LaGanga, 2015). A second approach would be to investigate the factors associated with patient no-shows from information attributed to them, allowing improvements in management practice and developing predictive models to avoid a potential no-show (Dantas et al., 2019, 2018; Griffin, 1998). These approaches can be applied with alternatives to improve attendance, such as call or confirmation messages or even a fee if the patient misses the appointment (Johnson et al., 2007).

Therefore, this work aims to understand which variables most influence patients' non-attendance in a cardiology and neurology clinic from Rio de Janeiro, Brazil, and to develop a prediction model to estimate the patient probability of no-show. The results can enable actions by managers focused on minimizing no-shows and improving scheduling policy. The present work is divided into the following sections. Section 2 presents the material and methods, Section 3 presents the main results, and Section 4 the discussion. Finally, the conclusion is presented in Section 5.

2 MATERIALS AND METHODS

The methodology is based on a case study to evaluate the factors related to no-shows in a clinic from Resende, Rio de Janeiro. This clinic offers different services, such as medical consultations and health procedures in the cardiological and neurology areas. The service is provided for all ages, between Monday and Friday, except for holidays, from 6 am to 8 pm, for patients with or without insurance.

We used deidentified data from January 2019 to June 2022, consisting of 32,573 registries. Initially, we decided to exclude some registries, as follows: personal slots of the physicians, which occurs when the doctor needs to close their schedule to other personal activities (outside the clinic); appointments previously canceled by patients since we are only considering no-shows without canceling; appointments on the weekend because of the low frequency; appointments of specific physicians no longer working in the clinic; and appointments with missing data in the binary response variable “show/no-show”. The outcome variable, no-show, is defined as a patient failure to attend the appointment without canceling at least 48 hours before the appointment time. The unit of analysis is the appointment record, not the patient.

Aiming to use the variables according to the need of this work and seeking a better understanding of the factors that impact the no-show, we performed a feature engineering. The “Insurance Company” variable details whether the patient pays directly to the clinic or uses private insurance plans. We divided this factor into three groups: the five more representative insurance companies in terms of the number of appointments (“5 Major”), other minor companies (“Minor”), and particular patients (“no insurance”). Following the same idea, the variable “Specialty” was aggregated into five categories: “Cardiology Appointment”, “Neurology Appointment”, “Procedure - Echocardiogram / Carotid, “Electroencephalography / Polysomnography Procedure”, and “Map / Holter Procedure”. The eleven variables used in the study and their description can be summarized in Table S1 (Supplementary Material I). A descriptive analysis was developed to explore the variables present in the database - the categorical data were expressed as frequency and proportion.

The appointment data points were randomly divided into development/training and validation/testing cohorts (80% and 20% of the data, respectively), which provides an unbiased sense of model effectiveness. Firstly, we used the training set to perform Cramér’s V statistics collinearity analysis and check for correlation between the included features (Kuhn and Johnson, 2013). Then, a simple logistical regression was performed to investigate the isolated effect of each explanatory variable on the dependent variable (no-show). Variables with a p-value less than $\alpha = 0.25$ were considered candidates for inclusion in the first multivariable model (Multivariable I).

Multivariable logistic regressions were performed to obtain the impact of each variable on patient no-shows adjusted by the other features. We also developed a second multivariable model (Multivariable II), in which the significance level of $\alpha = 0.05$ was used to select the statistically significant variables. The effective measures evaluated were the Odds Ratio (OR) and the p-value. The OR represents the relative effect that a factor level has on another one (called reference). The p-value informs us of the significance of that explanatory variable.

We used the Multivariable II prediction model to obtain the patient risk of being a no-show. We evaluated our final model based on discrimination and calibration measures. Discrimination was measured by the Area Under the Curve (AUC), and calibration was addressed by the calibration belts (Walsh et al., 2017). Still, other indicators were analyzed, such as the Positive Predictive Value (PPV), Negative Predictive Value (NPV), Sensitivity, and Specificity. The best cut-off for the final model’s prediction was determined based on the Youden index statistics. The complete information about the prediction methodology is summarized in Supplementary Material II.

All analyses were performed using R version 3.6.3, and the developed codes, model and dataset are publicly available on GitHub (https://github.com/igor-peres/No-show-prediction-models).

3 RESULTS

3.1 Data Preparation

The initial dataset presented 32,573 registries. After preparation and the data exclusion processes, 17,218 registries remained for data analysis and statistical modeling. Figure 1 shows the detailed exclusion process. The general no-show rate was 7.5% (1,300 patients).

Table 1 presents the descriptive analysis of each variable. It can be observed that the age groups between 21 and 50 years have the highest proportion of patients who missed the consultations, especially the age group from 31 to 40 years (10%). The high number of no-shows in August (8.4%) and December (12%) is also evident. Regarding the variable "Insurance company", we can observe a lower no-show for patients with minor representative insurance.

Table 1 – Descriptive analysis.

Independent Variables	No-show N (%)	Show N (%)	Total
Total	1,300 (7.5%)	15,918 (92.5%)	17,218
Age Group			
0 to 10 years	92 (5.4%)	1,627 (94.6%)	1,719
11 to 20 years	118 (7.7%)	1,414 (92.3%)	1,532
21 to 30 years	161 (9.7%)	1,505 (90.3%)	1,666
31 to 40 years	316 (10%)	2,722 (90%)	3,038
41 to 50 years	312 (9.2%)	3,071 (90.8%)	3,383
51 to 60 years	132 (5.8%)	2,130 (94.2%)	2,262
61 to 70 years	84 (5.0%)	1,608 (95%)	1,692
Over 70 years	85 (4.4%)	1,841 (95.6%)	1,926
Insurance company			
5 Major	1,011 (8.3%)	11,179 (91.7%)	12,190
Minors	135 (5.1%)	2,509 (94.9%)	2,644
Particular (no insurance)	154 (6.5%)	2,230 (93.5%)	2,384
Month			
January	117 (7.8%)	1,382 (92.2%)	1,499
February	119 (6.9%)	1,602 (93.1%)	1,721
March	129 (8.4%)	1,402 (91.6%)	1,531
April	100 (7.1%)	1,307 (92.9%)	1,407
May	106 (6.9%)	1,432 (93.1%)	1,538
June	112 (7.9%)	1,306 (92.1%)	1,418
July	81 (5.5%)	1,389 (94.5%)	1,470

Continued on next page

Table 1 – continued from previous page

Independent Variables	**No-show N (%)**	**Show N (%)**	**Total**
August	125 (8.4%)	1,365 (91.6%)	1,490
September	82 (6.4%)	1,193 (93.6%)	1,275
October	107 (6.9%)	1,434 (93.1%)	1,541
November	93 (7.3%)	1,173 (92.7%)	1,266
December	129 (12%)	933 (88%)	1,062
Time of the Day			
Before 9 am	131 (6.7%)	1,830 (93.3%)	1,961
From 9 am to 12 pm	372 (7.0%)	4,924 (93%)	5,296
From 12 pm to 3 pm	278 (7.2%)	3,593 (92.8%)	3,871
From 3 pm to 6 pm	471 (9.2%)	4,662 (90.8%)	5,133
After 6 pm	48 (5.0%)	909 (95%)	957
Day of the Week			
Monday	418 (8.0%)	4,790 (92%)	5,208
Tuesday	229 (9.2%)	2,269 (90.8%)	2,498
Wednesday	218 (6.8%)	2,980 (93.2%)	3,198
Thursday	269 (6.8%)	3,665 (93.2%)	3,934
Friday	166 (7.0%)	2,214 (93%)	2,380
Sex			
Female	799 (8.4%)	8,726 (91.6%)	9,525
Male	501 (6.5%)	7,192 (93.5%)	7,693
Scheduling Status			
Confirmed	942 (8.7%)	9,873 (91.3%)	10,815
Unconfirmed	358 (5.6%)	6,045 (94.4%)	6,403
Patient Type			
First Time	1,206 (7.7%)	14,449 (92.3%)	15,655
Return	94 (6.0%)	1,469 (94%)	1,563
Specialty			
Consultation – Cardiologist	447 (7.6%)	5,468 (92.4%)	5,915
Consultation – Neurologist	611 (9.8%)	5,597 (90.2%)	6,208
Procedure - Echocardiogram/Carotid	64 (3.8%)	1,622 (96.2%)	1,686
Procedure - Electroencephalogram/Polysomnography	138 (4.8%)	2,749 (95.2%)	2,887
Procedure - Map/Holter	40 (7.7%)	482 (92.3%)	522

According to the days of the week, we can note that during the beginning of the week, on Mondays and Tuesdays, patients tend to miss more appointments, reaching 8% and 9.2% of no-shows, respectively. It was also observed that the hours outside the commercial period, early morning and late afternoon, present the lowest no-show rates. It was not noted any effect influencing the

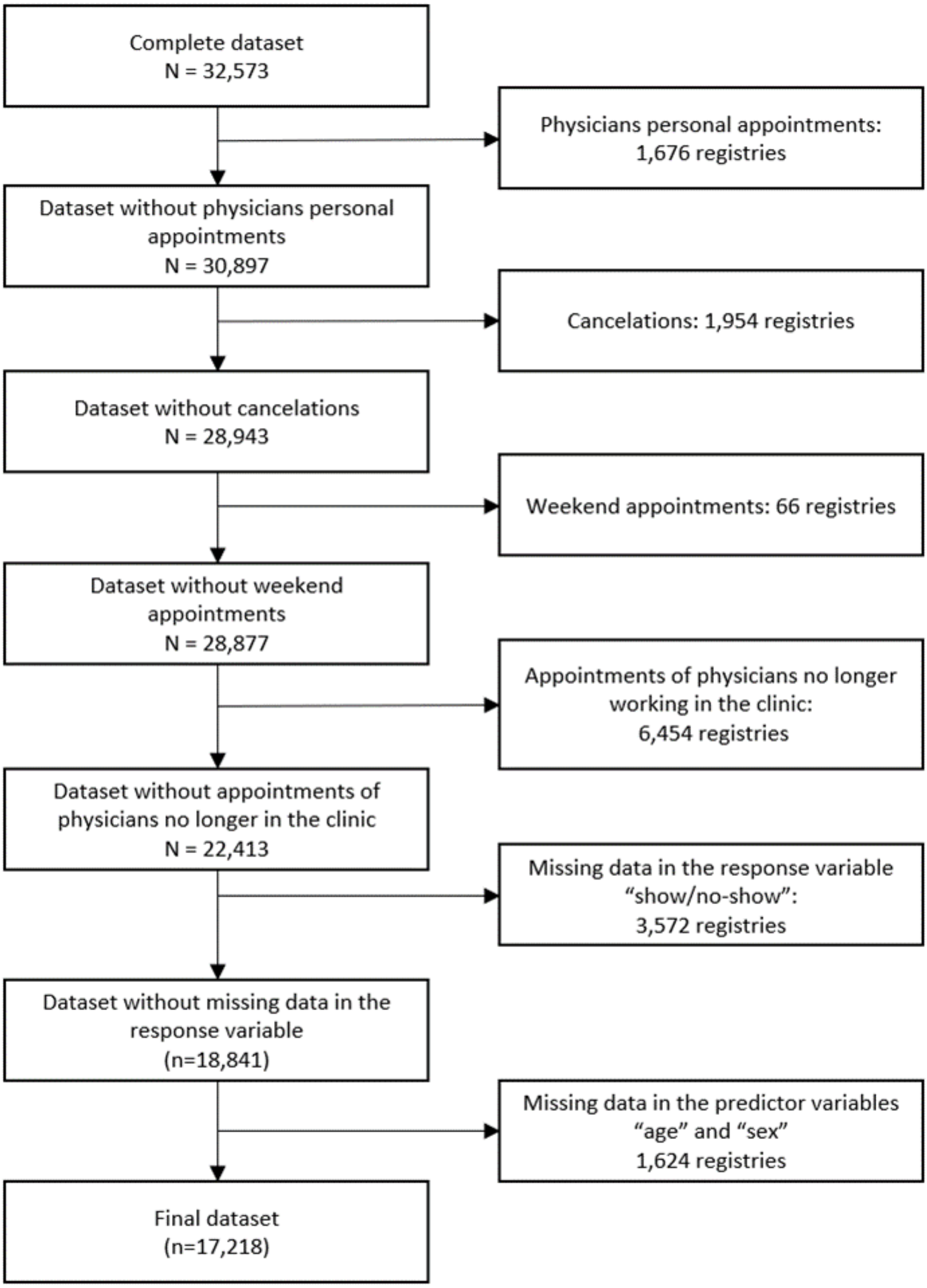

Figure 1 – Process of study inclusion.

no-show rate between the years. Other findings are that women are more representative than men and present the highest proportion of no-shows. Moreover, there is a greater concentration of appointments of first-time patients compared to returns and more no-shows in appointments than procedures.

Figure 2 shows the collinearity analysis of the included features. The greatest correlation was 0.33 between "Time of the Day" and "Specialty". It shows us that no significant collinearity justifies a possible feature elimination. Regarding the correlation between the independent variables with the dependent one (No-show), we also noted low correlations, being "Specialty" (0.11) and "Age" (0.08) the most correlated features.

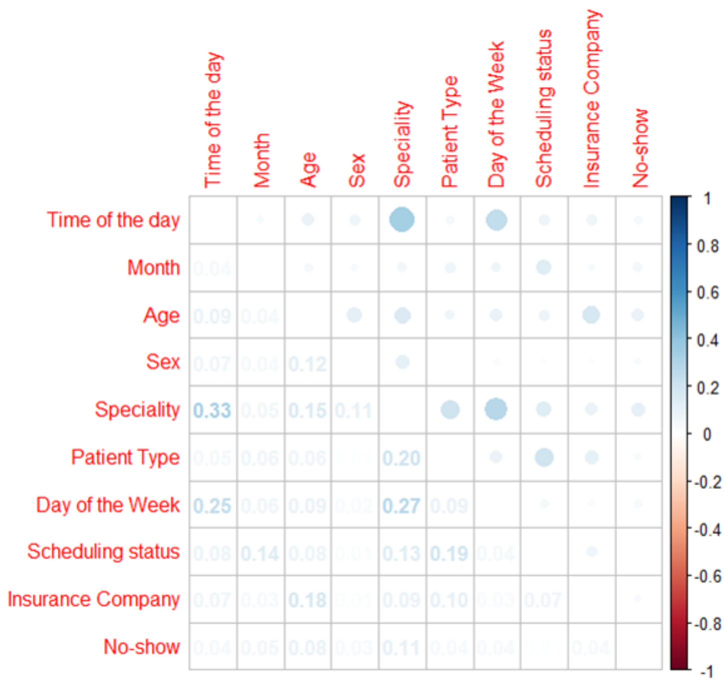

Figure 2 – Collinearity analysis of the features.

3.2 Logistic Regression

This subsection presents the univariate and multivariable analysis results. The univariate logistic regression aims to know which variables influenced the no-show rate, one at a time, and the multivariable regression identifies each variable's effect on the no-show rate when adjusted to the other variables in the model. The models can predict the patient risk of a no-show.

Table 2 shows coefficients, ORs, OR (with 95% confidence interval), and p values for each independent variable and model. The reference level for each category is indicated in parentheses for each variable. In this case, OR $>$ 1 predicts a lower likelihood of showing up for an appointment when compared to the reference level. In contrast, OR < 1 indicates a greater chance of showing up for an appointment.

According to the simple logistic regression, December has the greatest OR (1.61, CI = 1.28-2.03) compared to the reference (January). The variable sex is also significant, with a chance of a woman's non-attendance OR equal to 1.27 (1.14-1.40) compared to a man. We can observe that the variable "Patient Type" is significantly related to no-shows, with a first-time patient having a OR equal to 1.69 (1.38-2.09) compared to the return patients. Patients between 21 and 50 years of age presented the higher ORs, especially in the age group between 31 and 40 years, which shows a no-show twice as high compared to the age group from 0 to 10 years of age (OR = 2.03, CI = 1.61-2.58). All classes of the "Insurance company" were considered significant, which "no insurance" and "5 Major" presenting an OR equal to 1.35 (1.12-1.63) and 1.43 (1.24-1.67) compared to patients from "Minor" insurance companies, respectively. The variable "Specialty" is also significant for all categories. In this case, an OR of 2.36 (1.85-3.05) is observed for cardiac and 3.51 (2.77-4.52) for neurology appointments compared to the "Echocardiogram / Carotid procedure". The variables "Day of the Week" and "Time of the Day" were also significant in the univariable analysis.

We then developed a multivariable regression (Multivariable I). We identified that certain variables did not maintain significant p-value, meaning that some variables lost significance after adjusting for other confounding features. It occurred in "Day of the Week", "Time of the Day", and "Patient Type", where the p-value were all greater than 0.05. The OR values for the variables "Sex", "Insurance company", and "Month" did not vary too much when compared to the univariable regression. In addition, the age categories "21 to 30 years", "31 to 40 years", and "41 to 50 years" remain significant, although with lower odds ratio when compared to the univariable regression. Cardiology and neurology consultation and time of the day also remained significant in the multivariable analysis. Then, we performed a second multivariable logistic regression (Multivariable II), excluding the non-significant features of Multivariable I ("Day of the Week", "Time of the Day", and "Patient Type").

From final model (Multivariable II), we can see that: appointments scheduled in July presented lower no-show probability (OR = 0.65, CI= 0.47-0.91, p-value = 0.013) and in December higher chance (OR = 1.53, CI= 1.14-2.07, p-value = 0.005) compared to January; female sex presented more no-show compared to male (OR = 1.17, CI = 1.02-1.33, p-value = 0.023); ages categories "21 to 30 years", "31 to 40 years", and "41 to 50 years" presented greater risk to be a no-show (OR = 1.52, CI = 1.13-2.05, p-value = 0.006; OR = 1.58, CI = 1.21-2.08, p-value < 0.001; OR = 1.46, CI = 1.12-1.92, p-value= 0.006; respectively) compared to "0 to 10 years"; patients who have an insurance from "5 Major" companies tend to have a higher non-attendance (OR = 1.47, CI = 1.19-1.82, p-value < 0.001) compared to the smallest insurances; and slots scheduled to appointments to neurologist and cardiologist presented greater no-show (OR = 2.07, CI = 1.55-2.82, p-value < 0.001; OR = 2.41, CI = 1.81-3.27, p-value < 0.001; respectively) compared to echocardiogram/carotid procedure.

Table 2 – Univariable and Multivariable analyses.

Independent variable (reference level)	**Univariable**			**Multivariable I**			**Multivariable II**			
	OR	**IC-95%**	**p-value**	**OR**	**IC-95%**	**p-value**	β	**OR**	**IC-95%**	**p-value**
Intercept					-3.664					
Scheduling status (not confirmed)			0.4							
Confirmed	1.04	0.94-1.15	0.4							
Month (January)			**<0.001**			**<0.001**			**<0.001**	
February	0.88	0.70-1.11	0.30	0.86	0.66-1.13	0.3	-0.186	0.83	0.61-1.12	0.20
March	1.04	0.83-1.31	0.70	1.04	0.80-1.36	0.70	0.015	1.01	0.76-1.36	>0.90
April	0.91	0.72-1.16	0.50	0.88	0.67-1.17	0.40	-0.123	0.88	0.65-1.21	0.40
May	0.96	0.76-1.22	0.80	0.84	0.64-1.11	0.20	-0.168	0.85	0.62-1.14	0.30
June	0.88	0.69-1.13	0.40	1.02	0.78-1.34	0.90	0.018	1.02	0.75-1.38	>0.90
July	0.81	0.63-1.03	0.084	0.70	0.52-0.94	0.02	-0.425	0.65	0.47-0.91	**0.013**
August	1.11	0.89-1.40	0.40	1.04	0.80-1.35	0.80	0.013	1.01	0.75-1.36	>0.90
September	0.86	0.67-1.11	0.20	0.81	0.60-1.09	0.20	-0.171	0.84	0.61-1.16	0.30
October	0.97	0.77-1.22	0.80	0.90	0.68-1.19	0.50	-0.159	0.85	0.63-1.16	0.30
November	0.97	0.76-1.23	0.80	0.92	0.69-1.22	0.60	-0.112	0.89	0.65-1.23	0.50
December	1.61	1.28-2.03	<0.001	1.60	1.23-2.10	<0.001	0.426	1.53	1.14-2.07	**0.005**
Sex (Male)			**<0.001**			**0.018**			**0.023**	
Female	1.27	1.14-1.40	<0.001	1.17	1.03-1.34	0.018	0.153	1.17	1.02-1.33	**0.023**
Patient Type (Return)			**<0.001**			0.40				
First Time	1.69	1.38-2.09	<0.001	1.11	0.86-1.44	0.40				
Day of the Week (Monday)			**<0.001**			0.10				
Tuesday	0.95	0.82-1.10	0.50	1.16	0.95-1.43	0.14				
Wednesday	0.73	0.63-0.85	<0.001	1.23	0.99-1.53	0.059				
Thursday	0.78	0.68-0.89	<0.001	0.98	0.81-1.18	0.80				
Friday	0.76	0.65-0.89	<0.001	0.92	0.72-1.17	0.50				
Age (0 to 10 years)			**<0.001**			**<0.001**			**<0.001**	
11 to 20 years	1.44	1.09-1.91	0.009	1.31	0.98-1.75	0.071	0.270	1.31	0.96-1.79	0.091
21 to 30 years	1.84	1.42-2.40	<0.001	1.62	1.23-2.14	<0.001	0.418	1.52	1.13-2.05	**0.006**
31 to 40 years	2.03	1.61-2.58	<0.001	1.75	1.36-2.26	<0.001	0.457	1.58	1.21-2.08	**<0.001**
41 to 50 years	1.75	1.39-2.23	<0.001	1.60	1.25-2.07	<0.001	0.375	1.46	1.12-1.92	**0.006**
51 to 60 years	1.07	0.82-1.40	0.60	0.99	0.75-1.32	>0.90	-0.129	0.88	0.65-1.20	0.40

Continued on next page

Table 2 – continued from previous page

Independent variable (reference level)	**Univariable**			**Multivariable I**			**Multivariable II**			
	OR	**IC-95%**	**p-value**	**OR**	**IC-95%**	**p-value**	β	**OR**	**IC-95%**	**p-value**
61 to 70 years	0.92	0.69-1.24	0.60	0.85	0.62-1.16	0.30	-0.260	0.77	0.55-1.09	0.14
Over 70 years	0.80	0.60-1.07	0.14	0.74	0.54-1.02	0.065	-0.428	0.65	0.46-0.92	**0.014**
Time of the day (Before 9 am)			**<0.001**			0.40				
From 9 am to 12 pm	1.17	0.98-1.40	0.09	1.09	0.84-1.44	0.50				
From 12 pm to 3 pm	1.10	0.91-1.33	0.30	1.02	0.77-1.37	0.90				
From 3 pm to 6 pm	1.37	1.15-1.64	<0.001	1.20	0.91-1.59	0.20				
After 6 pm	0.75	0.56-1.0	0.049	0.99	0.65-1.48	>0.90				
Insurance Company (Minor)			**<0.001**			**<0.001**			**<0.001**	
Particular "no insurance"	1.35	1.12-1.63	0.002	1.36	1.07-1.74	0.014	0.239	1.27	0.97-1.67	0.086
5 Major	1.43	1.24-1.67	<0.001	1.46	1.22-1.77	<0.001	0.382	1.47	1.19-1.82	**<0.001**
Specialty (Procedure - Echocardiogram/Carotid)			**<0.001**			**<0.001**			**<0.001**	
Procedure - Electroencephalogram/Polysomnography	1.46	1.12-1.94	0.007	1.35	0.97-1.88	0.075	0.135	1.14	0.82-1.62	0.40
Procedure - Map/Holter	2.03	1.37-2.97	<0.001	2.25	1.41-3.57	<0.001	0.650	1.92	1.20-3.02	**0.006**
Appointment - Cardiologist	2.36	1.85-3.05	<0.001	2.22	1.65-3.01	<0.001	0.729	2.07	1.55-2.82	**<0.001**
Appointment - Neurologist	3.51	2.77-4.52	<0.001	2.50	1.90-3.34	<0.001	0.880	2.41	1.81-3.27	**<0.001**

Next, we used the Multivariable II model to predict the patient risk as a no-show in the testing set (validation part of the final dataset). The optimal cut-off found was 0.074. This small value can be explained due to an imbalanced database between the two groups (show and no-show). The results in terms of discrimination were: AUC = 0.65, Sensitivity = 0.64, Specificity = 0.58, PPV = 0.11, and NPV = 0.95 (using the optimal cut-off). Figure 3 shows the calibration belt for the predicted probabilities obtained in the testing set. We can observe a calibrated curve and small values for predicted probabilities, explaining the low value for the optimal cut-off. A practical application of the predictive no-show model can be seen in Supplementary Material II.

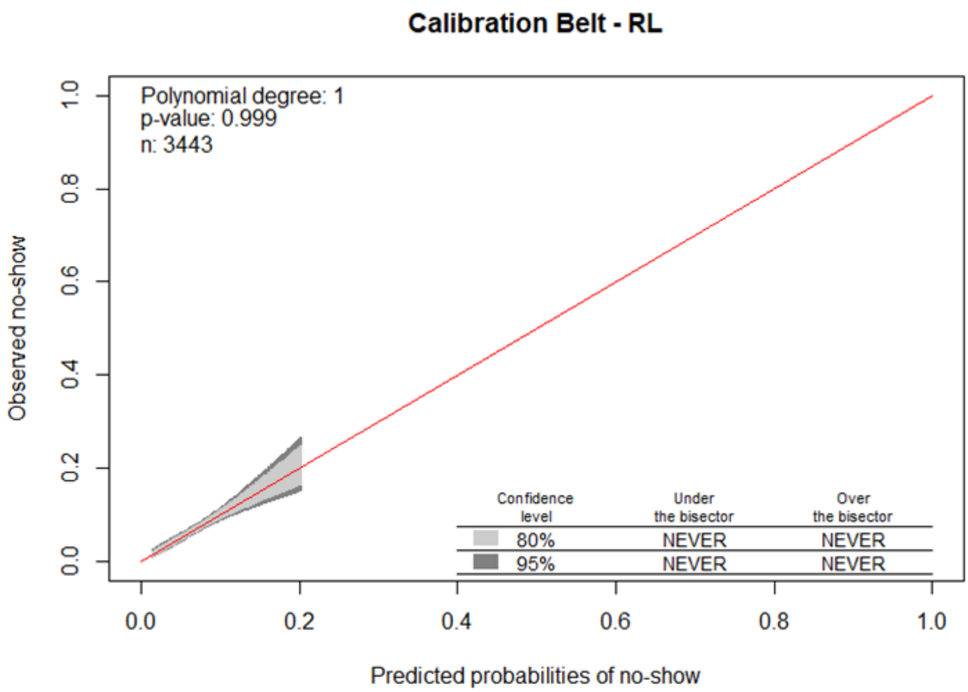

Figure 3 – Calibration belt for the predicted probabilities.

4 DISCUSSION

In this study, we analyzed the factors related to no-shows in a clinic from Rio de Janeiro. We have a no-show rate of 9.5%, a value lower than that found in other health clinics (mean of 23%, as noted by Dantas et al. (2018)). Furthermore, previous studies indicated 30% of no-shows in cardiac consultations and about 42% in neurology consultations (Dantas et al., 2018), values far from those found in this study (9.2% and 13%, respectively). However, how the no-shows are defined and quantified varies from study to study and could justify the differences between the no-show rate of each literature study case. As an example, it is worth saying that some studies usually account for canceled consultations, among other occurrences (Milicevic et al., 2020).

A second hypothesis for this difference may be related to the type of service provided in the clinic since most studies consider unique appointments. Another hypothesis is that these clinic appointments are linked to medical procedures, which have a lower rate of no-shows due to the difficulty of getting an appointment. Finally, it also can be related to the clinic process of confirmation and follow-ups, which tends to reduce the no-show rates.

The most significant features in our study were sex, age, insurance company, and specialty. It is notorious that women tend to seek more medical care when compared to men (CNN Brasil, 2022). In this case study, this difference is evidenced when we observed that women represent about 10% more than the male group in clinical care. It relates to the hypothesis that men choose to go to a medical consultation when their cases are more critical (UFMG Faculty of Medicine, 2021). The number of no-shows for women in this work was 10%, while for men, it was 8.4%. These values are also similarly found in related articles (Torres et al., 2015). Studies also indicate that patient age is related to non-attendance. We showed that the clinic patients with greater OR mainly were people aged 21- 50 (able to join the economically active population). We hypothesize that they often have to miss appointments due to work. Although insignificant, 3 pm to 6 pm demonstrates a chance of missing an appointment about 1.21 times compared to the reference (9 am). It also happens in previous studies in which it was reported that there is a higher number of no-shows in the afternoon (Agarwal et al., 2022; Elkhider et al., 2022; Torres et al., 2015). When analyzing the insurance company, the patient with less representative insurance may have more difficulty scheduling their appointment and may miss fewer appointments, which may explain the greater no-show for individuals with major insurance. We also proposed a multivariable model to predict the patient risk to be a no-show. The model's AUC was equal to 0.65, which is acceptable for this problem. Moreover, using the optimal cut-off, the proposed model showed a sensitivity of 0.64, implying a 64% capacity to identify the positive no-show cases. Identifying possible no-show patients is crucial to prioritize the confirmation process and to optimize scheduling policies.

Our work has the following limitations. First, we did not analyze the repeated measures correlation for return appointments since we did not have the patient identification. Second, we did not include other possible features because they were unavailable, such as distance to home, weather, and traffic on the appointment day. Third, we noted that the appointment confirmation, which the literature commonly finds as significantly associated with non-attendance, appeared not to be related to a no-show in our analysis (p-value = 0.4), and it may have occurred because of the lack of confirmation report by the secretaries.

5 CONCLUSIONS

This work evaluated the effect of demographic and clinical variables on patient no-shows through a case study in a clinic from Rio de Janeiro, Brazil. It showed that female patients, adults, with major insurance plans, and those scheduled for an appointment (and not a procedure) have a higher risk of being a possible no-show. Such conclusions have the potential to assist in understanding and improving the efficiency of the clinic's appointment scheduling policy. In this sense, some suggestions can be made, such as elaborating a confirmation strategy to consider these conclusions about factors most related to non-attendance. For future work, it is desired to validate our predictive model in other contexts. We also want to propose new scheduling policies considering the patient no-show probability generated by our model (and evaluate it through simulation).

Acknowledgements

The data that support the findings of this study are openly available in GitHub at https://github.com/igor-peres/No-show-prediction-models. The authors received no financial support for the research, authorship, and publication of this article.

References

AGARWAL P, NATHAN A, JALEEL Z & LEVI J. 2022. Factors Contributing to Missed Appointments in a Pediatric Otolaryngology Clinic. *The Laryngoscope*, **132**: 895–900.

BHATTACHARJEE P & RAY P. 2016. Simulation modelling and analysis of appointment system performance for multiple classes of patients in a hospital: A case study. *Operations Research for Health Care*, **8**: 71–84.

CAYIRLI T & VERAL E. 2003. Outpatient Scheduling in Health Care: A Review of Literature. *Production and Operations Management*, **12**: 519–549.

CNN BRASIL. 2022. CNN Brasil. Available at: https://www.cnnbrasil.com.br/saude/mais-de-60-dos-homens-brasileiros-so-procuram-medico-com-sintomas-insuportaveis/.

COELHO E, MOYSÉS NETO M, PALHARES R, CARDOSO MM, GELEILETE T & NOBRE F. 2005. Relação entre a assiduidade às consultas ambulatoriais e o controle da pressão arterial em pacientes hipertensos. *Arq. Bras. Cardiol*, **85**: 157–161.

DANTAS L, FLECK J, CYRINO OLIVEIRA F & HAMACHER S. 2018. No-shows in appointment scheduling – a systematic literature review. *Health Policy*, **122**: 412–421.

DANTAS L, HAMACHER S, CYRINO OLIVEIRA F, BARBOSA S & VIEGAS F. 2019. Predicting Patient No-show Behavior: a Study in a Bariatric Clinic. *OBES SURG*, **29**: 40–47.

ELKHIDER H, SHARMA R, SHENG S, THOSTENSON J, KAPOOR N, VEERAPANENI P, SIDDAMREDDY S, IBRAHIM F, YADALA S, ONTEDDU S & NALLEBALLE K. 2022. Predictors of No-Show in Neurology Clinics. *Healthcare*, **10**: 599. Available at: https://doi.org/10.3390/healthcare10040599.

GOFFMAN R, HARRIS S, MAY J, MILICEVIC A, MONTE R, MYASKOVSKY L, RODRIGUEZ K, TJADER Y & VARGAS D. 2017. Modeling Patient No-Show History and Predicting Future Outpatient Appointment Behavior in the Veterans Health Administration. *Military Medicine*, **182**: 1708– 1714. Available at: https://doi.org/10.7205/MILMED-D-16-00345.

GRIFFIN S. 1998. Lost to follow-up: the problem of defaulters from diabetes clinics. *Diabetic Medicine*, **15**: 14– 24. Available at: https://doi.org/10.1002/(SICI)1096-9136(1998110)15:3+.

GUPTA D & WANG WY. 2012. Patient Appointments in Ambulatory Care. In: HALL R (Ed.), *Handbook of Healthcare System Scheduling, International Series in Operations Research & Management Science*. p. 65–104. Boston, MA: Springer US.

JOHNSON B, MOLD J & PONTIOUS J. 2007. Reduction and Management of No-Shows by Family Medicine Residency Practice Exemplars. *The Annals of Family Medicine*, **5**: 534–539.

KIM JH. 2009. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational Statistics & Data Analysis*, **53**: 3735–3745.

KOPACH R, DELAURENTIS PC, LAWLEY M, MUTHURAMAN K, OZSEN L, RARDIN R, WAN H, INTREVADO P, QU X & WILLIS D. 2007. Effects of clinical characteristics on successful open access scheduling. *Health Care Manage Sci*, **10**: 111–124.

KUHN M & JOHNSON K. 2013. Data Pre-processing. In: KUHN M & JOHNSON K (Eds.), *Applied Predictive Modeling*. p. 27–59. New York, NY: Springer.

LAGANGA L & LAWRENCE S. 2007. Clinic Overbooking to Improve Patient Access and Increase Provider Productivity. *Decision Sciences*, **38**: 251–276.

MILICEVIC A, MITSANTISUK K, TJADER A, VARGAS D, HUBERT T & SCOTT B. 2020. Modeling Patient No-Show History and Predicting Future Appointment Behavior at the Veterans Administration's Outpatient Mental Health Clinics: NIRMO-2. *Military Medicine*, **185**: 988–994.

PARIKH A, GUPTA K, WILSON A, FIELDS K, COSGROVE N & KOSTIS J. 2010. The Effectiveness of Outpatient Appointment Reminder Systems in Reducing No-Show Rates. *The American Journal of Medicine*, **123**: 542–548.

PERES I, HAMACHER S, CYRINO OLIVEIRA F, BARBOSA S & VIEGAS F. 2019. Simulation of Appointment Scheduling Policies: a Study in a Bariatric Clinic. *OBES SURG*, **29**: 2824–2830.

RUFIBACH K. 2010. Use of Brier score to assess binary predictions. *Journal of Clinical Epidemiology*, **63**: 938–939.

SAMORANI M & LAGANGA L. 2015. Outpatient appointment scheduling given individual day-dependent no-show predictions. *European Journal of Operational Research*, **240**: 245–257.

TOPUZ K, UNER H, OZTEKIN A & YILDIRIM M. 2018. Predicting pediatric clinic no-shows: a decision analytic framework using elastic net and Bayesian belief network. *Ann Oper Res*, **263**: 479–499.

TORRES O, ROTHBERG M, GARB J, OGUNNEYE O, ONYEMA J & HIGGINS T. 2015. Risk Factor Model to Predict a Missed Clinic Appointment in an Urban, Academic, and Underserved Setting. *Population Health Management*, **18**: 131–136.

TUSO P, MURTISHAW K & TADROS W. 1999. The Easy Access Program: A Way to Reduce Patient No-Show Rate, Decrease Add-Ons to Primary Care Schedules, and Improve Patient Satisfaction. *The Permanente Journal*, **3**: 68–71. Available at: https://doi.org/10.7812/TPP/99.951.

UFMG FACULTY OF MEDICINE. 2021. Available at: https://www.medicina.ufmg.br/homens-ainda-estao-muito-atras-das-mulheres-nos-cuidados-com-a-saude/.

WALSH C, SHARMAN K & HRIPCSAK G. 2017. Beyond discrimination: A comparison of calibration methods and clinical usefulness of predictive models of readmission risk. *Journal of Biomedical Informatics*, **76**: 9–18.

How to cite

PERES IT, DANTAS LF, GOUVEIA DM & SCHNEIDER F. 2023. Prediction and factors related to no-show: a study in a cardiology and neurology clinic. *Pesquisa Operacional*, **43**: e275795. doi: 10.1590/0101-7438.2023.043.00275795.

APPENDIX A SUPPLEMENTARY MATERIAL I

Table A1 – Characterization of variables used in this study.

Variable	**Description**	**Type**
Service Status	Show/No-Show	Categorical
Year	Year of patient scheduling	Categorical
Day of the Week	Patient scheduling day	Categorical
Specialty	Type of schedule offered by the clinic	Categorical
Timetable	Patient scheduling time	Categorical
Age	Patient age, divided into age ranges from 10 to 10 years	Categorical
Month	Patient scheduling month	Categorical
Insurance Company	Payment method chosen by the patient	Categorical
Sex	Classification of patients' sex by name according to IBGE census 2010	Categorical
Scheduling Status	If the patient confirms the consultation	Categorical
Patient Type	Patient's first time in the office or if he is returning	Categorical

APPENDIX B SUPPLEMENTARY MATERIAL II

B.1 Statistical methods

The appointment data points were randomly divided into training and testing cohorts (containing 80% and 20% of the data, respectively). The training cohort was used to create the logistic regression model to estimate a patient's no-show probability and the testing cohort to find the model's accuracy, comparing the expected number to the actual number of no-shows.

Then, a univariable regression was performed to investigate the isolated effect of each explanatory variable on the dependent variable (no-show), and the multivariable logistic regression to obtain the impact of each variable on patient no-shows adjusted by the other features.

From the multivariable logistic regression model, we can estimate the no-show probability for each individual by feeding the model with the values of predictive variables associated with the particular patient. Equation (B.1) indicates a general form of the logistic regression model.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta \cdot X \tag{B.1}$$

where p is the no-show probability, β_0 is a constant, X is the matrix of independent variables for predicting the no-show risk, and β is the vector of coefficients corresponding to X, representing the relationship between the variables and the reference level. The term p/p-1 is known as the odds of the missed appointment risk (Odds Ratio - OR). In this article, matrix X comprises the factors “Month”, “Sex”, “Age”, “Patient Type”, “Insurance company”, “Day of the week”, “Time of the day”, and “Specialty”.

B.2 Practical application of the predictive no-show model

Based on the full logistic regression model (Equation B.1), one can estimate the no-show probability for each individual by feeding it with the values of the independent variables associated with a particular patient.

To demonstrate the practical use of the predictive model from Table 2, the full logistic regression model for the no-show probability is as follows (Equation B.2):

$$\ln\left(\frac{\widehat{p}}{1-\widehat{p}}\right) = \widehat{\beta_0} + \widehat{\beta_1}X1 + \widehat{\beta_2}X2 + \widehat{\beta_4}X4 + \widehat{\beta_5}X5 + \widehat{\beta_6}X6 \tag{B.2}$$

The coefficient (β) of the level selected for each factor X is imputed and calculated. The coefficient is zero whether the chosen level is the reference level.

Let us consider an example. A 28-year-old male patient is scheduled to see the Cardiologist for an appointment “Particular”. This appointment will be on Tuesday, in June, at 3 pm. Using the coefficients from Table 2, the calculation of the no-show rate for this patient using the Multivariate II model is the following (Equation 3):

$$\ln\frac{\widehat{p}}{1-\widehat{p}} = -3.664 + 0.018 + 0.2390 + 0.418 + 0 + 0.729 = -2.26$$

$$\frac{\widehat{p}}{1-\widehat{p}} = e^{-2.26} = 0.1043 \rightarrow \widehat{p} = 0.1043\,(1-\widehat{p}) \rightarrow \widehat{\mathbf{p}} = \mathbf{0.094} \tag{B.3}$$

It results in a probability of a 9.4% chance of missing his next appointment. Comparing this probability with the choice cut-off (0.074), this patient would be in the no-show class. The results of this empirical study can be used to change scheduling policies, such as overbooking and developing a dynamic scheduling system that considers each patient’s no-show probability.