

Modeling the No-Show of Patients to Exam Appointments of Computed Tomography

Rodolfo Benedito Zattar Silva (✉ rodolfoengenheiro@hotmail.com)

Universidade Federal do Rio Grande do Sul <https://orcid.org/0000-0003-2075-5622>

Flávio Sanson Fogliatto

Universidade Federal do Rio Grande do Sul

Tiago Severo Garcia

Hospital de Clinicas de Porto Alegre

Carlo Sasso Faccin

Hospital de Clinicas de Porto Alegre

Arturo Alejandro Zavala Zavala

Universidade Federal de Mato Grosso

Research article

Keywords: No-show modeling, No-show predictors, Radiology, Computed tomography, Variable selection.

Posted Date: September 30th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-70869/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: No-shows of patients have negative impacts on healthcare systems, such as resources' underutilization, efficiency loss, and cost increase. Predicting no-show is key to develop strategies that counteract its effects. In this paper, we propose a model to predict the no-show of ambulatory patients to exam appointments of computed tomography at the Radiology department of a large Brazilian public hospital.

Methods: We carried out a retrospective study on 8,382 appointments made to computed tomography (CT) exams between January and December 2017. Penalized logistic regression and multivariate logistic regression were used to model the influence of 15 candidate variables on patients' no-show. The predictive capabilities of the models were evaluated analyzing the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC).

Results: The no-show rate in computerized tomography exams appointments was 6.65%. The two models performed similarly in terms of AUC. The penalized logistic regression model was selected using the parsimony criterion, with 8 of the 15 variables analyzed appearing as significant. One of the variables included in the model (number of exams scheduled in previous year) had not been previously reported in the related literature.

Conclusions: Our findings may be used to guide the development of strategies to reduce the no-show of patients to exam appointments.

Background

A no-show occurs when a patient does not show up for a scheduled appointment or does not cancel the appointment such that the time slot can be assigned to another patient. The occurrence of no-shows is one of the main operational problems found in outpatient scheduling processes [1]. No-shows have negative impacts on most healthcare systems, such as reduced efficiency, increased costs, lost revenues, and limited access of patients to services [2]. Patient no-show to radiology exams lead to even greater losses due to the high cost of imaging equipment and professionals required for capturing and interpreting images [3].

No-show rates may vary from 3–80% in healthcare systems, depending on the service being provided and patients' demographic characteristics [4]. Studies report different no-show rates; e.g., 18.80% in a study carried out in a medical center offering consultation on 10 different ambulatory specialties [4], 21.90% in bariatric surgery appointments in a specialized clinic - considering only consultations before and after surgery [2], and 6.50% in a radiology department considering all exams performed [3]. In the Brazilian public health network, patient no-show has become a chronic problem, similar to what is observed in other countries. No-show rates to consultations and exams are observed as high in different regions of the country and across medical specialties [5].

There is consensus in the literature regarding the non-random nature of no-shows, and research has been published proposing statistical analyses of factors that may impact no-show rates [6, 7]. Previous studies have shown that factors associated with no-show include patients' sociodemographic conditions such as age [4]; gender [7, 8]; distance between home and healthcare unit [9, 10]; race [11]; marital status [12, 13]; level of schooling [14]; appointment characteristics – time [2], day of the week [9], and month [15]; elapsed time between appointment request and actual consultation [2]; and patient's clinical diagnosis [16].

In a recent literature review article [6], 105 works that analyzed the influence of 45 different variables in patient no-show were identified. However, none of these studies analyzed the association between the predictor variable "number of exams scheduled in the previous year" and no-show. Such variable describes patients' past behavior regarding exam appointments and is not likely to be directly available in consultation databases. In addition, there were no studies analyzing factors associated with the no-show in a Brazilian radiology department. Regionalizing no-show studies is relevant since it is well known that sociodemographic conditions impact patients' consultations and exams attendance ratios. There are two contributions here. The first contribution regards the state-of-the-art of no-show modeling: the study of a new predictor associated with the no-show may contribute to future analyses of the phenomenon in other healthcare settings. The second contribution regards the state-of-the-practice on no-show management: the use of a database from a Brazilian radiology department, which allows managers to identify characteristics of patients and service that determine non-attendance to image exams. Considering the scarcity of resources in the Brazilian public health system, it is critical for developing strategies to reduce no-show and mitigate its negative effects on the system; such actions may reduce the underutilization of services and reduce waiting times to obtain appointments [5].

The objective of the present study is to identify variables associated to patient no-show to exam appointments and build a prediction model that allows determining the no-show probability of future patients using historical data from the Radiology Department of a large Brazilian hospital.

Methods

Data, sample and predictive variables

We performed a retrospective study using outpatient computed tomography (CT) scan scheduling data obtained from the consultation scheduling management system of the Radiology Department of a public, 850-bed, tertiary care teaching hospital in southern Brazil. The study was approved by the Research Ethics Committee under project number CAEE 83645318.6.0000.5327.

The CT unit operates continuously (24/7), issuing approximately 30,000 reports per year, covering emergency patients, inpatients and outpatients referred to the hospital. We chose to focus our analysis on CT appointments for two reasons. First, a non-utilized CT slot costs more to the system than X-ray or ultrasound unused slots. Second, the number of CT exams in this department is larger than that of

magnetic resonance imaging (MRI) exams; in addition, a larger no-show rate is associated with CT exams (7.20%) than with MRI exams (5.10%), considering 2016 data. The unit has three CT equipment; 74 ambulatory patients are scheduled daily to perform exams, from 8AM to 5PM.

We considered in the study appointment records from patients that visited the hospital at least once in 2016 for medical consultations, laboratory or radiology exams. In 2017, 12,387 appointments were made for CT exams; 2,966 did not meet our first screening selection criteria. The 9,421 remaining appointment records covered a one-year period, from Jan 1 to Dec 31, 2017, being mostly abdominal (5,148) and skull and brain (1,051) CT exams. In the final sample analyzed we removed patients with more than one CT appointment on the same day, which led to 8,382 appointments (and a total of 557 no-shows) (see Fig. 1). Our dataset did not include appointments with missing information on the analyzed predictors.

Figure 1 Flow-chart of patient selection process

The dataset carries information on 15 categorical and continuous predictors, which were grouped into three main categories (see Additional file 1: Table S1). Variables included in the dataset were chosen based on the literature and availability of data in the scheduling system. The response (dependent) variable was defined to be the outpatient attendance status with two possible outcomes: show or no-show. The primary outcome (no-show) was associated with a patient that did not show to perform the scheduled CT or did not cancel the appointment.

Statistical analysis

We started by analyzing the sample's descriptive statistics. Absolute and relative frequencies were used to describe categorical variables, while continuous variables were described through measures of central tendency (mean) and dispersion (standard deviation—SD). Characteristics of patients in the show and no-show classes were compared through Chi-Squared and *t* tests, depending on the variable's type.

We analyzed no-show frequencies in ten percentiles of each continuous variable in Table S1 (Additional file 1). The objective was to visually identify the nature of the relationship between descriptive variables and patient no-show (linear, quadratic or logarithmic).

Next, the dataset was split into training and test portions. Records in each portion were randomly selected, however keeping the same proportion of no-shows in the complete sample (6.65%). The training portion was comprised of 5,867 records (403 no-shows) and used to develop the models; the test portion was comprised of 2,515 records (154 no-shows) and used to validate the models. Model performance was assessed determining the ROC-AUC for training and test portions.

To assess the generalizability of models tested, we used the Monte Carlo cross validation method (cross validation with repeated random sub-sampling) with 500 repetitions, using different portions of data for training and test (keeping the proportion of the positive class – no-show – in the two portions). Mean and standard deviation of 500 repetitions were calculated for AUC, sensitivity, specificity, Positive Predictive

Value (PPV) and Negative Predictive Value (NPV). Here, PPV refers to the probability of no-show to the CT exam, while the NPV refers to the probability of show to the CT exam.

Logistic regression with LASSO penalization (Least Absolute Shrinkage and Selection Operator) and logistic regression were used to analyze the factors associated with patient's no-show in the training portion of the dataset. Penalized logistic regression is a technique similar to logistic regression, although used for automatic selection of variables. For that, it imposes a penalty on the model β coefficients to induce dispersion, such that coefficients of less relevant predictors are reduced to zero [17]. This process results in biased coefficient estimates that can no longer be interpreted as odds ratios, as in a logistic regression, but rather as a weight or relative importance of the predictor variable [18]. The best value for the penalty parameter λ is the one that leads to the smallest mean square error (MSE) in the cross validation.

In the logistic regression, a univariate analysis was initially carried out to verify the influence of each predictive variable on the no-show. Variables with significance levels smaller than 0.25 ($P < 0.25$) were considered for inclusion in the multivariate analysis, following recommendations in previous studies [2, 16]. In the multivariate logistic regression analysis, a backward (upstream) variable selection was carried out, moving from an initial model including all variables retained in the univariate analysis, which were eliminated stepwise if displaying P values larger than 0.05. The final reduced model obtained comprised of only statistically significant variables ($P < 0.05$).

The objective was to choose a model to calculate patients' no-show probabilities to scheduled CT exam appointments. Using the test portion of the dataset the selected model goodness-of-fit was assessed using the Hosmer–Lemeshow test, which divides observations into classes based on their predicted no-show probabilities and calculates the chi-squared value associated with observed and expected frequencies in each class [19]. In addition, the ROC curve was used to determine the threshold probability value that best classifies patients (maximizing sensitivity and specificity). The predicted no-show probability of each patient was then compared to the threshold and those displaying larger probability values were classified as no-show. Finally, we generated the confusion matrix for the selected model using the test portion. All analyses were performed using the R package (version 3.5.1).

Results

In our sample of 8,382 patients scheduled for CT examinations in 2017, 7,825 (93.35%) attended and 557 (6.65%) did not attend the exams. According to Table 1, the highest proportion of no-shows took place among males (7.1%), black (10.3%), single (8.5%) patients with no schooling (10.1%) and no cancer (6.8%), with appointments on Thursdays (7.2%), in the afternoon (7.0%), in June (12.5%), and who had 3 or more no-shows to exams in the previous year (12.3%). Variables, such as age, distance, number of consultations scheduled, number of no-shows in consultations and number of exams scheduled, in previous year, presented average values slightly larger for no-show patients; whereas, lead time presented average value smaller for no-show patients. No significant differences in patients' characteristics in the

show and no-show classes were found for the following predictors: gender, cancer, day and shift (Table 1).

Table 1
Sample of patients – descriptive statistics

Predictors	Show (<i>N</i> = 7,825)	No-show (<i>N</i> = 557)	<i>P</i> value
Categorical			
Gender			
Female– <i>n</i> (%)	4,242 (93.7)	283 (6.3)	0.130 ^a
Male– <i>n</i> (%)	3,583 (92.9)	274 (7.1)	
Race			
Asian– <i>n</i> (%)	5 (100.0)	0 (0.0)	< 0.001 ^a
White– <i>n</i> (%)	6,998 (93.7)	467 (6.3)	
Native Brazilian– <i>n</i> (%)	1 (100.0)	0 (0.0)	
Brown– <i>n</i> (%)	248 (91.2)	24 (8.8)	
Black– <i>n</i> (%)	573 (89.7)	66 (10.3)	
Marital Status			
Single– <i>n</i> (%)	2,337 (91.5)	217 (8.5)	< 0.001 ^a
Married– <i>n</i> (%)	3,937 (94.4)	234 (5.6)	
Divorced– <i>n</i> (%)	822 (95.1)	42 (4.9)	
Widow(er)/Others– <i>n</i> (%)	729 (91.9)	64 (8.1)	
Schooling			
None– <i>n</i> (%)	543 (89.9)	61 (10.1)	0.011 ^a
Elementary incomplete– <i>n</i> (%)	3,783 (93.6)	260 (6.4)	
Elementary complete– <i>n</i> (%)	1,283 (93.4)	90 (6.6)	
High School incomplete– <i>n</i> (%)	407 (92.9)	31 (7.1)	
High School complete– <i>n</i> (%)	1,362 (93.6)	93 (6.4)	
^a Chi-square test, ^b <i>t</i> -test			

Predictors	Show (<i>N</i> = 7,825)	No-show (<i>N</i> = 557)	<i>P</i> value
Tertiary incomplete– <i>n</i> (%)	169 (97.1)	5 (2.9)	
Tertiary complete– <i>n</i> (%)	278 (94.2)	17 (5.8)	
Cancer			
No– <i>n</i> (%)	3,400 (93.2)	249 (6.8)	0.595 ^a
Yes– <i>n</i> (%)	4,425 (93.5)	308 (6.5)	
Day of the week			
Monday– <i>n</i> (%)	1,554 (92.9)	118 (7.1)	0.511 ^a
Tuesday– <i>n</i> (%)	1,610 (93.1)	119 (6.9)	
Wednesday– <i>n</i> (%)	1,592 (94.0)	102 (6.0)	
Thursday– <i>n</i> (%)	1,440 (92.8)	112 (7.2)	
Friday– <i>n</i> (%)	1,628 (93.9)	106 (6.1)	
Shift			
Morning– <i>n</i> (%)	3,756 (93.7)	252 (6.3)	0.224 ^a
Afternoon– <i>n</i> (%)	4,069 (93.0)	305 (7.0)	
Month			
January– <i>n</i> (%)	770 (93.3)	55 (6.7)	< 0.001 ^a
February– <i>n</i> (%)	578 (92.8)	45 (7.2)	
March– <i>n</i> (%)	696 (92.4)	57 (7.6)	
April– <i>n</i> (%)	503 (92.0)	44 (8.0)	
May– <i>n</i> (%)	624 (90.7)	64 (9.3)	
June– <i>n</i> (%)	559 (87.5)	80 (12.5)	
^a Chi-square test, ^b <i>t</i> -test			

Predictors	Show (<i>N</i> = 7,825)	No-show (<i>N</i> = 557)	<i>P</i> value
July– <i>n</i> (%)	637 (90.4)	68 (9.6)	
August– <i>n</i> (%)	703 (90.8)	71 (9.2)	
September– <i>n</i> (%)	573 (91.0)	57 (9.0)	
October– <i>n</i> (%)	646 (98.9)	7 (1.1)	
November– <i>n</i> (%)	784 (99.2)	6 (0.8)	
December– <i>n</i> (%)	752 (99.6)	3 (0.4)	
Number of no-shows to exams in the previous year			
0– <i>n</i> (%)	6,773 (94.0)	435 (6.0)	< 0.001 ^a
1– <i>n</i> (%)	685 (90.5)	72 (9.5)	
2– <i>n</i> (%)	217 (88.2)	29 (11.8)	
3 or more– <i>n</i> (%)	150 (87.7)	21 (12.3)	
Continuous			
Age–mean (SD)	58.48 (15.86)	57.61 (17.48)	< 0.001 ^b
Distance–mean (SD)	57.87 (99.52)	53.97 (103.66)	< 0.001 ^b
Lead time–mean (SD)	88.96 (71.44)	95.40 (59.76)	< 0.001 ^b
Number of consultations scheduled in previous year–mean (SD)	10.21 (10.74)	10.12 (11.06)	< 0.001 ^b
Number of no-shows to consultations in previous year– mean (SD)	0.90 (1.64)	1.14 (1.95)	< 0.001 ^b
Number of exams scheduled in previous year–mean (SD)	6.11 (7.18)	5.84 (7.36)	< 0.001 ^b
^a Chi-square test, ^b <i>t</i> -test			

Table 1 here.

The analysis of the type of relationship between continuous descriptive variables and patient no-show indicated the following: quadratic for age; logarithmic for distance, lead time, and number of

consultations and exams scheduled in the previous year; and linear for number of no-shows to consultations in the previous year (see Additional file 2: Table S2). Variable “number of no-shows to exams in the previous year” did not display any of the behaviors above and was thus treated as categoric, despite its continuous nature.

Regression coefficients (β) of the penalized logistic regression model, as well as regression coefficients (β), odds ratios, 95% confidence intervals, and P values associated with each variable in the analysis of univariate and multivariate logistic regressions, determined using the training dataset, are shown in Table 2. In that table, variables with no β coefficient values for the penalized logistic regression model their coefficients zeroed by the LASSO penalty.

Table 2

Regression coefficients of the penalized logistic regression model and regression coefficients, odds ratio values, 95% confidence intervals (CI95%), and *P* values of the univariate and multivariate logistic regressions models

Predictors	Penalized logistic regression	Univariate logistic regression (retain predictors with <i>P</i> < 0.25)			Multivariate logistic regression – reduced model (retain predictors with <i>P</i> < 0.05)		
	β	β	Odds ratio (CI95%)	<i>P</i> value	β	Odds ratio (CI95%)	<i>P</i> value
Intercept	-3.10				-3.47		
Gender							
Female			Ref.			Ref.	
Male		0.14	1.15 (0.97–1.36)	0.120	0.21	1.23 (1.03–1.47)	0.026
Race							
White			Ref.			Ref.	
Not white	0.21	0.49	1.63 (1.29–2.07)	0.000	0.37	1.45 (1.13–1.86)	0.004
Marital Status							
Single			Ref.			Ref.	
Married	-0.12	-0.45	0.64 (0.53–0.78)	0.000	-0.33	0.72 (0.59–0.88)	0.001
Divorced/Separated	-0.09	-0.60	0.55 (0.39–0.77)	0.001	-0.54	0.58 (0.41–0.82)	0.002
Widow(er)/Others		-0.06	0.95 (0.71–1.26)	0.706	0.03	1.04 (0.76–1.41)	0.826
Schooling							
None			Ref.			Ref.	

Ref. reference group, Predictors with a *P* value no more than 0.25 in the univariate logistic regression analysis were included in the multivariate logistic regression analysis

Predictors	Penalized logistic regression	Univariate logistic regression (retain predictors with $P < 0.25$)			Multivariate logistic regression – reduced model (retain predictors with $P < 0.05$)		
	β	β	Odds ratio (CI95%)	P value	β	Odds ratio (CI95%)	P value
Elementary incomplete		-0.49	0.61 (0.46–0.82)	0.001	-0.38	0.68 (0.50–0.92)	0.014
Elementary complete		-0.47	0.62 (0.44–0.88)	0.007	-0.37	0.69 (0.48–0.98)	0.039
High School incomplete		-0.39	0.68 (0.43–1.06)	0.091	-0.38	0.68 (0.43–1.08)	0.106
High School complete		-0.50	0.61 (0.43–0.85)	0.004	-0.41	0.66 (0.47–0.94)	0.022
Tertiary incomplete		-1.33	0.26 (0.10–0.67)	0.005	-1.28	0.28 (0.11–0.71)	0.008
Tertiary complete		-0.61	0.54 (0.31–0.95)	0.032	-0.42	0.66 (0.37–1.17)	0.155
Cancer							
No			Ref.				
Yes		-0.05	0.95 (0.80–1.13)	0.564			
Day of the week							
Monday			Ref.				
Tuesday		-0.03	0.97 (0.75–1.27)	0.841			
Wednesday		-0.17	0.84 (0.64–1.11)	0.224			

Ref. reference group, Predictors with a P value no more than 0.25 in the univariate logistic regression analysis were included in the multivariate logistic regression analysis

Predictors	Penalized logistic regression	Univariate logistic regression (retain predictors with $P < 0.25$)			Multivariate logistic regression – reduced model (retain predictors with $P < 0.05$)		
	β	β	Odds ratio (CI95%)	P value	β	Odds ratio (CI95%)	P value
Thursday		0.03	1.02 (0.78–1.34)	0.861			
Friday		-0.15	0.86 (0.65–1.12)	0.265			
Shift							
Morning			Ref.				
Afternoon		0.11	1.12 (0.94–1.33)	0.208			
Month							
January			Ref.			Ref.	
February		0.09	1.09 (0.72–1.64)	0.679	0.10	1.10 (0.73–1.67)	0.640
March		0.14	1.15 (0.78–1.68)	0.486	0.11	1.11 (0.75–1.64)	0.590
April		0.20	1.22 (0.81–1.85)	0.335	0.17	1.18 (0.78–1.80)	0.428
May		0.36	1.44 (0.99–2.09)	0.059	0.40	1.49 (1.02–2.19)	0.039
June	0.34	0.69	2.00 (1.40–2.87)	0.000	0.71	2.03 (1.40–2.93)	0.000
July		0.40	1.49 (1.03–2.17)	0.034	0.34	1.41 (0.97–2.06)	0.073

Ref. reference group, Predictors with a P value no more than 0.25 in the univariate logistic regression analysis were included in the multivariate logistic regression analysis

Predictors	Penalized logistic regression	Univariate logistic regression (retain predictors with $P < 0.25$)			Multivariate logistic regression – reduced model (retain predictors with $P < 0.05$)		
	β	β	Odds ratio (CI95%)	P value	β	Odds ratio (CI95%)	P value
August		0.35	1.41 (0.98–2.04)	0.064	0.20	1.22 (0.84–1.77)	0.298
September		0.33	1.39 (0.95–2.05)	0.093	0.21	1.23 (0.83–1.82)	0.307
October	-1.00	-1.89	0.15 (0.07–0.34)	0.000	-2.02	0.13 (0.06–0.30)	0.000
November	-1.18	-2.23	0.11 (0.05–0.25)	0.000	-2.31	0.10 (0.04–0.23)	0.000
December	-1.31	-2.88	0.06 (0.02–0.18)	0.000	-2.97	0.05 (0.02–0.17)	0.000
Number of no-shows to exams in the previous year							
0			Ref.			Ref.	
1	0.17	0.49	1.64 (1.26–2.13)	0.000	0.59	1.81 (1.35–2.41)	0.000
2	0.30	0.73	2.08 (1.40–3.10)	0.000	0.90	2.46 (1.59–3.79)	0.000
3 or more	0.20	0.78	2.18 (1.37–3.48)	0.001	0.76	2.13 (1.26–3.59)	0.004
Age		-0.03	0.97 (0.99–1.00)	0.218			
Age ²		0.00	1.00 (1.00–1.00)	0.530			

Ref. reference group, Predictors with a P value no more than 0.25 in the univariate logistic regression analysis were included in the multivariate logistic regression analysis

Predictors	Penalized logistic regression	Univariate logistic regression (retain predictors with $P < 0.25$)			Multivariate logistic regression – reduced model (retain predictors with $P < 0.05$)		
	β	β	Odds ratio (CI95%)	P value	β	Odds ratio (CI95%)	P value
Log (Distance)	-0.02	-0.07	0.93 (0.89–0.97)	0.000	-0.08	0.93 (0.89–0.97)	0.000
Log (Lead time)	0.16	0.24	1.27 (1.14–1.41)	0.000	0.33	1.39 (1.23–1.56)	0.000
Log (Number of consultations scheduled in previous year)		-0.06	0.94 (0.88–0.99)	0.031	-0.09	0.91 (0.83–0.99)	0.046
Number of no-shows to consultations in previous year	0.03	0.07	1.08 (1.03–1.12)	0.001	0.09	1.10 (1.04–1.16)	0.001
Log (Number of exams scheduled in previous year)	-0.02	-0.04	0.96 (0.93–0.99)	0.009	-0.05	0.95 (0.91–0.99)	0.018
Ref. reference group, Predictors with a P value no more than 0.25 in the univariate logistic regression analysis were included in the multivariate logistic regression analysis							

Table 2 here.

The penalized logistic regression model kept eight significant no-show predictors. Fourteen predictors appeared as statistically significant in the univariate logistic regression and included in the multivariate, which retained 11 in the final model (Table 2).

Using results in Table 2, we verified that both penalized and multivariate logistic regression models displayed the same discrimination power between patients in show and no-show classes. No significant differences were found between AUC values for the penalized logistic regression and the multivariate logistic model in the training portion ($P = 0.907$; Fig. 2A) and test portions ($P = 0.949$; Fig. 2B). Both models also presented similar AUC, sensitivity, specificity, PPV and NPV values in the cross validation (Table 3). To select the best prediction model to patients' no-show in future CT appointments we applied the principle of parsimony (i.e. simpler models should be chosen over more complex ones), since the both models displayed similar performance [20].

Figure 2. (A) ROC curves for no-show prediction in the training portion; (B) ROC curves for no-show predictions in the test portion

Table 3
Classification metrics obtained through cross-validation

Metrics	Penalized logistic regression	Multivariate logistic regression
AUC	0.716 ± 0.014	0.720 ± 0.016
Sensitivity	0.728 ± 0.044	0.756 ± 0.090
Specificity	0.570 ± 0.025	0.573 ± 0.095
PPV	0.108 ± 0.009	0.115 ± 0.015
NPV	0.967 ± 0.005	0.971 ± 0.007
<i>AUC</i> area under the curve, <i>PPV</i> positive predictive value, <i>NPV</i> negative predictive value		

The eight predictors of no-show to CT exam appointments selected by the penalized logistic regression model are: race, marital status, month, number of no-shows to exams and consultations in the previous year, distance, lead-time and number of exams scheduled in previous year. Predictors with positive (negative) associated regression coefficients are indicators of higher (lower) no-show probability. Analyzing the categorical predictors, we conclude that patients with higher no-show probabilities are non-white, with appointments scheduled for June, with 1, 2, 3 or more no-shows to exams in the previous year. Patients with smaller no-show probabilities are those married or divorced and scheduled for the months of October, November and December (Table 2).

Analyzing results for the continuous predictors (Table 2), we verify that the no-show probability decreases with the increase in distance values and in the number of exams scheduled in the previous year. In opposition, no-show probability increases with the rise in lead time and number of no-shows in consultations in the previous year. We present as Additional file an example of practical application of the chosen prediction model using information of a hypothetical patient (see Additional file 3).

The Hosmer–Lemeshow test validated the penalized logistic regression model applied to the test portion of the dataset, i.e., the difference between the observed and predicted probabilities was not significant ($P = 0.135$). Observed and predicted probabilities to each decile of the test portion are presented in Table 4. The cut-off point in the ROC curve for the test portion was 0.0645 or (6.45%). Using that threshold value, we obtained the confusion matrix displayed in Table 5. The chosen prediction model presented a sensitivity of 0.786, specificity of 0.616, PPV of 0.118 and NPV of 0.978.

Table 4
Observed and predicted probabilities for each decile in the test portion

Decile	Observed probabilities	Predicted probabilities
1	0.000	0.006
2	0.008	0.013
3	0.012	0.024
4	0.024	0.037
5	0.048	0.049
6	0.052	0.060
7	0.067	0.072
8	0.080	0.086
9	0.100	0.106
10	0.222	0.161

Table 5
Confusion matrix for the penalized logistic regression model in the test portion

Real	Predicted		
	No-show	Show	Total
No-show	True positives (TP): 121	False positives (FP): 33	154
Show	False negatives (FN): 906	True negatives (TN): 1,455	2,361
Total	1,027	1,488	

Discussion

The no-show rate to CT exams found in this study (6.65%) is close to those observed in studies carried out using radiology department datasets from countries such as the United States – 7.70% [3] and 7.50% [21] - and Australia – 6.50% [15], which reinforces no-show to exams of CT as a global problem and a challenge to be faced by healthcare systems' managers.

No-show predictors appearing as significant in this study were also observed in previous works. Other studies also identified non-white patients as more likely to not show up [3, 9]. In the penalized logistic regression model, the predictor with largest regression coefficient (and therefore largest contribution to no-show) was the month of June. That may be related to the fact that in the Southern hemisphere winter

begins in June. This effect is consistent with that observed in a Danish radiology department, in which winter was associated with larger no-show probabilities [22].

The month for which exam is scheduled also appeared as a risk factor in other studies dealing with no-show [14, 23]. In a study conducted at a university hospital in South America [23], it also was found that patients scheduled for the month of June were more likely not to show up at consultations in opposition to consultations scheduled for November, in which patients were found more likely to show up. However, two results do not agree with our findings; namely: (i) patients scheduled for the month of December were more likely not to show up; (ii) October was not significant month for no-show prediction [23].

The predictor number of no-shows to exams in the previous year is positively related to no-show, indicating that past no-show behavior is carried onto future appointments, as observed in previous studies [3, 24]. Likewise, the number of no-shows in consultations in the previous year was found statistically significant and positively associated to no-show in previous works [13, 15, 25].

Previous studies also negatively related the no-show probability with distance traveled by patients to reach the healthcare unit [2, 23, 26]. In our model, there are two possible explanations for this variable's behavior: first, most appointments made for the year 2017 (58%) are for patients from other cities for which free transportation is provided by their municipalities; second, patients lack other options of specialized radiology services in their hometowns. Other studies also positively related the no-show probability with the lead time between scheduling a medical appointment and getting it [2, 7, 11].

The present study contributes to the statistical modeling of no-show by investigating a new predictor, number of exams scheduled in the previous year. The predictor appeared significant and negatively correlated to the probability of no-show; i.e., the larger the number of exams scheduled in the previous year, the smaller the probability of no-show in new appointments; thereby agreeing with the perceptions of radiology department managers, who argue that patients requiring a more rigorous monitoring of healthcare treatments usually schedule more exams and are less likely to not show up.

On the other hand, predictors identified as non-significant in our work displayed similar behavior in previous studies; more specifically, gender [2, 27]; schooling [28]; cancer [15]; day [2, 16, 23, 25]; shift of the appointment [14, 16]; age [2, 11]; and number of consultations scheduled in previous year [29].

This work contributes to the existing literature on no-show modeling in Radiology [3, 9, 15, 21, 22, 24, 30–33] in several ways. First, by corroborating findings on no-show predictors in previous works and compiling such results (see Additional file 4: Table S3). Second, previous studies only investigated the use of traditional logistic regression. Here, we extend the analysis to include the penalized logistic regression which proved to be a less complex modeling alternative, more robust to new data and with results that are more easily interpreted and applied in practice. The penalized logistic regression had been previously applied in the modeling of no-show [17, 18] but not in the context of a Radiology department.

The use of a predictive model for patient no-show as presented in this work has implications to the allocation of CT resources. A high sensitivity model presents less false positives (patients wrongly classified as show) but higher proportion of false negatives (patients wrongly classified as no-show). That allows reducing the idleness in the use of CT resources, but overloads the system as many patients classified as no-show would appear for exams. A high specificity model, on the other hand, would not overload the system, but promote idleness in CT resources. The determination of the cut-off point in the ROC graph allows balancing the relationship between sensitivity and specificity, which was the course of action adopted here.

The findings in this study are also relevant to managers of radiology services in public institutions since the ability to predict patient no-shows to exams allows the planning of counteractions that improve results and minimize lost revenues [3]. Data released by the World Health Organization [34] show that in Brazil, public spending on health as a percentage of the Gross Domestic Product (GDP) remained practically stable (below 10%) between the years 2000 and 2017 (Additional file 4: Figure S1) and smaller than other countries such as France, which also has a universal and public health system. Low investment in public health limits capacity expansion penalizing users of the system, especially in the case of diagnostic imaging installations that have high fixed costs related to equipment, and the need for specialized staff to operate them. Thus, adopting measures to reduce no-show rates is one of the main strategies to increase access to exams. That is a critical step towards improving the quality of health care, especially considering that CT exams allows a more effective planning of care to patients, indicates the need for surgical interventions, reduces exploratory surgeries, improves the monitoring and treatment of cancer, guides the treatment of common injuries (injury, stroke, heart disease), and reduce the length of hospital stay [35].

Among possible actions to reduce the no-show of patients to CT exams in the analyzed department, we list the utilization of overbooking and reaching patients through text messages in the days leading up to the exam appointment. The use of text message reminders (Short Message Service - SMS) proved to be an effective strategy to decrease the no-show rate in a Radiology department [36]. However, the authors sent SMS notifications to all patients with scheduled appointments. Using the model presented in this work, it is possible to intensify the communication with patients more likely not to attend to the exams. The use of social media may also be considered for sending reminders, promoting awareness campaigns on the importance of exams and the impact of no-shows on the system. In Brazil, approximately 71% of the population are Internet users, of which 66% have active social media profiles [37].

We finalize by pointing some limitations in our study. First, results cannot be generalized to other healthcare units or regions in the country, since the dataset analyzed was obtained from a single radiology department of a public hospital located in the south of Brazil. Second, improvements in the predictive power of the model obtained here are constrained by data availability. Future research directions include the integration of results obtained in this study to exams scheduling systems and the analysis of counteractive initiatives to mitigate the impacts of patients' no-show in the radiology department analyzed here.

Conclusions

This study demonstrated that penalized logistic regression can be used to predict outpatient no-show to CT scan exams, considering information related to patients' sociodemographic and clinical conditions as well as scheduling characteristics.

We verified that patients with higher no-show probabilities are non-white, scheduled for June, with 1, 2, 3, or more no-shows in exams scheduled in the previous year, who reside closer to the study unit, present the smallest number of exams scheduled in the previous year, with larger lead times and number of no-shows in consultations in the previous year.

Results obtained here may subsidize strategies to mitigate the negative impacts of patients' no-show and improve the efficiency of operations in radiology departments.

Abbreviations

CT: Computed tomography; AUC: Area under the curve; ROC: Receiver operating characteristic; MRI: magnetic resonance imaging; SD: Standard-deviation; PPV: Positive predictive value; NPV: Negative predictive value; LASSO: Least absolute shrinkage and selection operator; MSE: Mean square error; GDP: Gross domestic product; SMS: Short message service.

Declarations

Acknowledgements

Not applicable.

Authors' contributions

RS contributed to study design, analysis of data and draft of the manuscript. FF contributed to the design of the study, drafting the article, and revising it critically for important academic content. TG and CF contributed to the acquisition, analysis, and interpretation of data. AZ contributed to study design, computational programming, analysis, and interpretation of data. All authors approved the final version for publication.

Funding

Not applicable.

Availability of data and materials

Based on a mutual agreement between researchers and Radiology Department of Hospital de Clínicas de Porto Alegre (HCPA), the datasets used during the present study are not publicly available.

Ethics approval and consent to participate

The study was approved by the Ethics Committee of Research of the Hospital de Clínicas de Porto Alegre (HCPA) under project number CAEE 83645318.6.0000.5327.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflict of interests.

References

1. Topuz K, Uner H, Oztekin A, Yildirim MB. Predicting pediatric clinic no-shows: a decision analytic framework using elastic net and Bayesian belief network. *Ann Oper Res* 2018; 263: 479–499.
2. Dantas LF, Hamacher S, Oliveira FLC, Barbosa SDJ, Viegas F. Predicting patient no-show behavior: a study in a bariatric clinic. *Obes Surg* 2019; 29: 40–47.
3. Harvey HB, Liu C, Ai J, Jaworsky C, Guerrier CE, Flores E, et al. Predicting no-shows in radiology using regression modeling of data available in the electronic medical record. *J Am Coll Radiol* 2017; 14:1303–1309.
4. Kheirkhah P, Feng Q, Travis LM, Tavakoli-Tabasi S, Sharafkhaneh A. Prevalence, predictors and economic consequences of no-shows. *BMC Health Serv Res* 2016; 16: 13.
5. Oleskovicz M, Oliva FL, Grisi CC de H e, Lima AC, Custódio I. Overbooking in an outpatient healthcare facility in the Brazilian Unified National Health System. *Cad Saúde Pública* 2014; 30: 1009-1017.
6. Dantas LF, Fleck JL, Cyrino Oliveira FL, Hamacher S. No-shows in appointment scheduling - a systematic literature review. *Health Policy* 2018.
7. Peng Y, Erdem E, Shi J, Masek C, Woodbridge P. Large-scale assessment of missed opportunity risks in a complex hospital setting. *Informatics Heal Soc Care* 2016; 41: 112–27.
8. Lenzi H, Ben ÂJ, Stein AT. Development and validation of a patient no-show predictive model at a primary care setting in Southern Brazil. *PLoS One* 2019; 14: e0214869.
9. Daye D, Carrodegua E, Glover M, Guerrier CE, Harvey HB, Flores EJ. Impact of delayed time to advanced imaging on missed appointments across different demographic and socioeconomic factors. *J Am Coll Radiol* 2018;15: 713–720.
10. Goffman RM, Harris SL, May JH, Milicevic AS, Monte RJ, Myaskovsky L, et al. Modeling patient no-show history and predicting future outpatient appointment behavior in the veterans health administration. *Mil Med* 2017;182: e1708–14.
11. Lotfi V, Torres E. Improving an outpatient clinic utilization using decision analysis-based patient scheduling. *Socioecon Plann Sci* 2014; 48: 115–126.

12. Blumenthal DM, Singal G, Mangla SS, Macklin EA, Chung DC. Predicting non-adherence with outpatient colonoscopy using a novel electronic tool that measures prior non-adherence. *J Gen Intern Med* 2015; 30: 724–731.
13. Dobbs RW, Malhotra NR, Caldwell BM, Rojas R, Moreira DM, Abern MR. Determinants of clinic absenteeism: a novel method of examining distance from clinic and transportation. *J Community Health* 2018; 43: 19–26.
14. Huang Y-L, Hanauer DA. Time dependent patient no-show predictive modelling development. *Int J Health Care Qual Assur* 2016; 29: 475–488.
15. Mander GTW, Reynolds L, Cook A, Kwan MM. Factors associated with appointment non-attendance at a medical imaging department in regional Australia: a retrospective cohort analysis. *J Med Radiat Sci* 2018; 192-199.
16. Daggy J, Lawley M, Willis D, Thayer D, Suelzer C, DeLaurentis P-C, et al. Using no-show modeling to improve clinic performance. *Health Informatics J* 2010;16: 246–259.
17. Lin Q, Betancourt B, Goldstein BA, Steorts RC. Prediction of appointment no-shows using electronic health records. *J Appl Stat* 2019; 1–15.
18. Ding X, Gellad ZF, Mather C, Barth P, Poon EG, Newman M, et al. Designing risk prediction models for ambulatory no-shows across different specialties and clinics. *J Am Med Informatics Assoc* 2018; 25: 924–930.
19. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. New Jersey: John Wiley & Sons, Inc.; 2013.
20. Pernía-Espinoza A, Fernandez-Ceniceros J, Antonanzas J, Urraca R, Martinez-de-Pison FJ. Stacking ensemble with parsimonious base models to improve generalization capability in the characterization of steel bolted components. *Appl Soft Comput J* 2018; 70: 737–750.
21. Glover M, Daye D, Khalilzadeh O, Pinykh O, Rosenthal DI, Brink JA, et al. Socioeconomic and demographic predictors of missed opportunities to provide advanced imaging services. *J Am Coll Radio* 2017; 14: 1403–1411.
22. Blaehr EE, Søgaaard R, Kristensen T, Væggemose U. Observational study identifies non-attendance characteristics in two hospital out-patient clinics. *Dan Med J* 2016; 63:1–6.
23. Giunta D, Briatore A, Baum A, Luna D, Waisman G, Gonzalez BQF. Factors associated with nonattendance at clinical medicine scheduled outpatient appointments in a university general hospital. *Patient Prefer Adherence* 2013; 7: 1163–1170.
24. Parente CA, Salvatore D, Gallo GM, Cipollini F. Using overbooking to manage no-shows in an Italian healthcare center. *BMC Health Serv Res* 2018; 18: 185.
25. Guzek LM, Gentry SD, Golomb MR. The estimated cost of “no-shows” in an academic pediatric neurology clinic. *Pediatr Neurol* 2015; 52:198–201.
26. Storrs MJ, Ramov HM, Lalloo R. An investigation into patient non-attendance and use of a short-message reminder system at a university dental clinic. *J Dent Educ* 2016; 80:30–39.

27. Huang Y, Hanauer DA. Patient no-show predictive model development using multiple data sources for an effective overbooking approach. *Appl Clin Inform*. 2014; 5: 836–860.
28. Cheng K-D, Huang C-J, Tsang H-Y, Lin C-H. Factors related to missed first appointments after discharge among patients with schizophrenia in Taiwan. *J Formos Med Assoc* 2014; 113: 436–441.
29. Coodin S, Staley D, Cortens B, Desrochers R, McLandress S. Patient factors associated with missed appointments in persons with schizophrenia. *Can J Psychiatry* 2004; 49:145–148.
30. Lu JC, Lowery R, Yu S, Ghadimi Mahani M, Agarwal PP, Dorfman AL. Predictors of missed appointments in patients referred for congenital or pediatric cardiac magnetic resonance. *Pediatr Radiol* 2017; 47: 911–916.
31. O. AlRowaili M, Ahmed AE, Areabi HA. Factors associated with No-Shows and rescheduling MRI appointments. *BMC Health Serv Res* 2016; 16: 1–7.
32. Mieloszyk RJ, Rosenbaum JI, Hall CS, Hippe DS, Gunn ML, Bhargava P. Environmental factors predictive of no-show visits in radiology: observations of three million outpatient imaging visits over 16 years. *J Am Coll Radiol* 2019; 16: 554–559.
33. Rosenbaum JI, Mieloszyk RJ, Hall CS, Hippe DS, Gunn ML, Bhargava P. Understanding why patients no-show: observations of 2.9 million outpatient imaging visits over 16 years. *J Am Coll Radiol* 2018; 15: 944–950.
34. OMS. Database, Global Health Expenditure.
<https://apps.who.int/nha/database/ViewData/Indicators/es>. Accessed March 31, 2020.
35. Kordolaimi SD, Efstathopoulos EP. Computed tomography radiation dosimetry: from the indicators to the indications. *J Comput Assist Tomogr* 2014; 38: 807–814.
36. Liu C, Harvey HB, Jaworsky C, Shore MT, Guerrier CE, Pinykh O. Text message reminders reduce outpatient radiology no-shows but do not improve arrival punctuality. *J Am Coll Radiol* 2017;1049–1054.
37. DataReportal. Digital 2020: global digital yearbook. <https://datareportal.com/reports/digital-2020-global-digital-yearbook>. Accessed January 28, 2020.

Figures

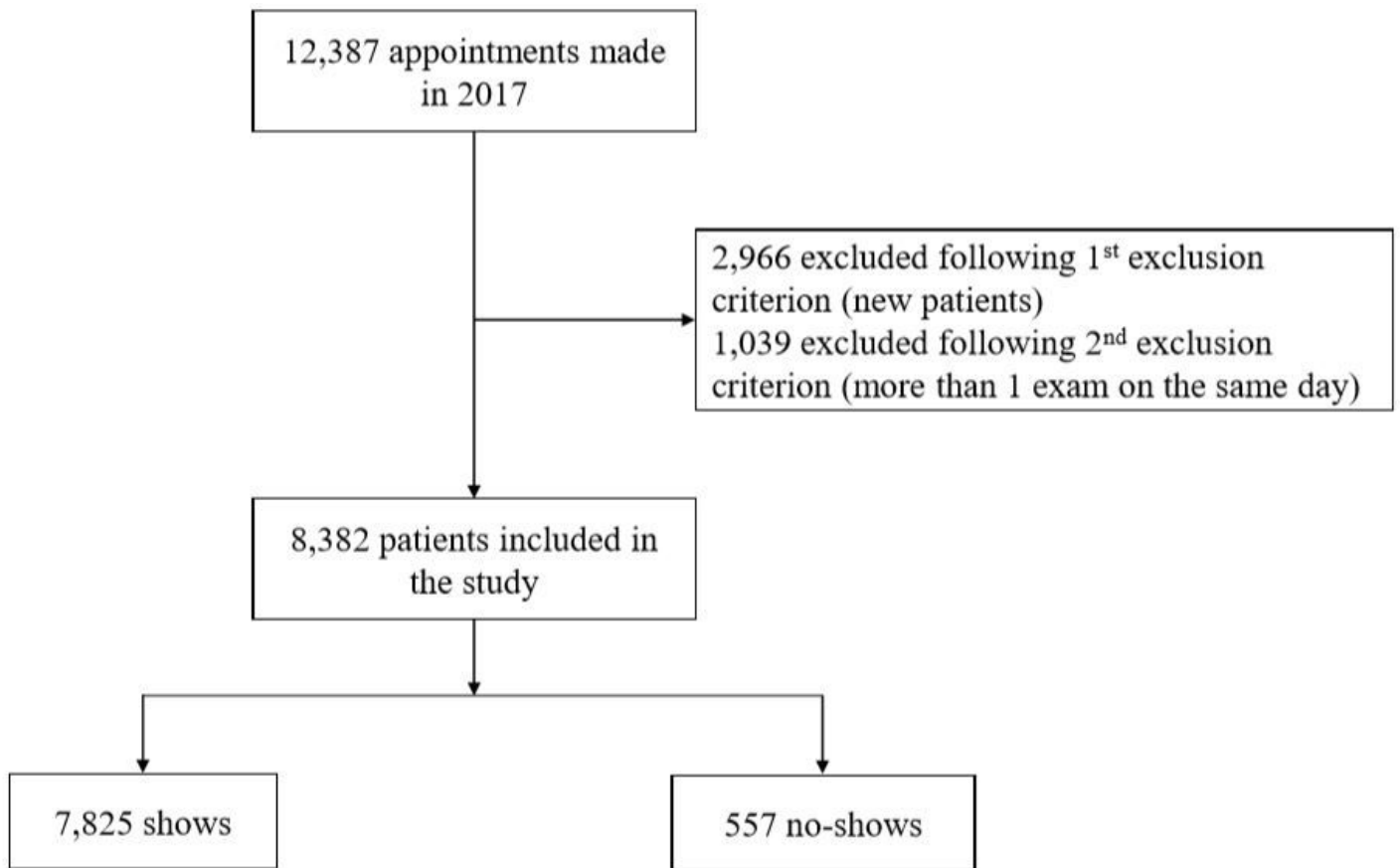


Figure 1

Flow-chart of patient selection process

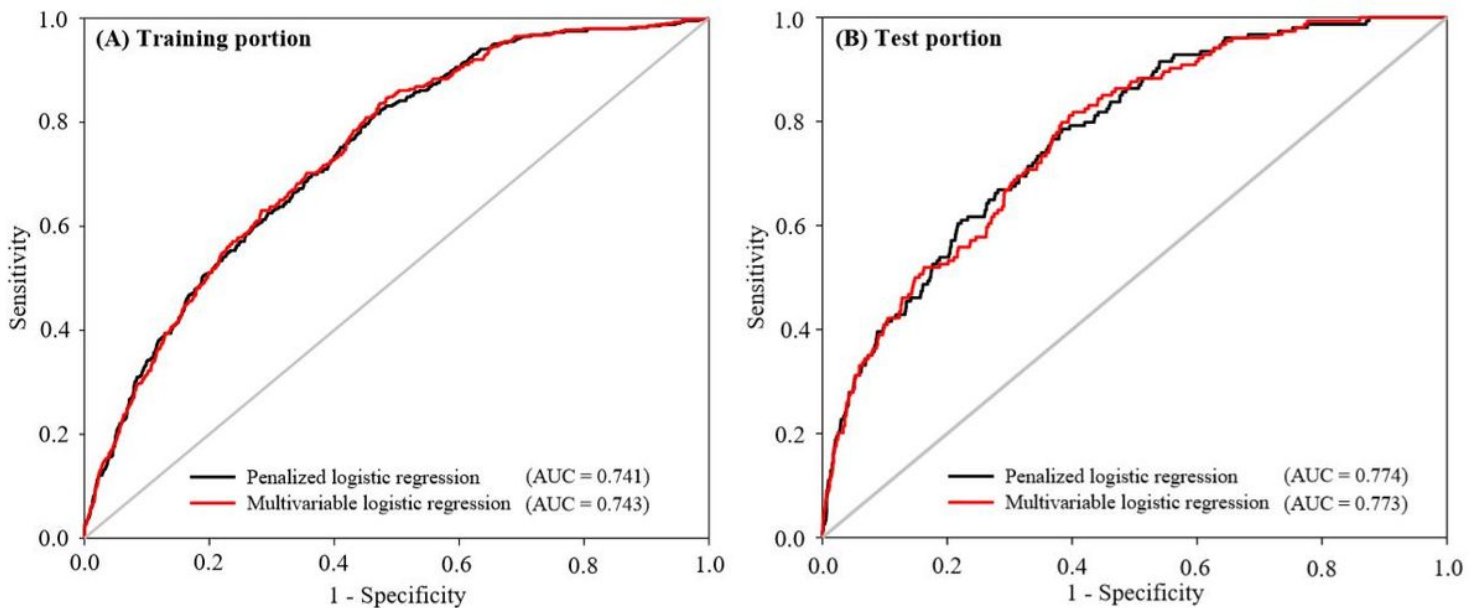


Figure 2

(A) ROC curves for no-show prediction in the training portion; (B) ROC curves for no-show predictions in the test portion

Supplementary Files

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

- [Additionalfile1.pdf](#)
- [Additionalfile2.pdf](#)
- [Additionalfile3.pdf](#)
- [Additionalfile4.pdf](#)
- [Additionalfile5.pdf](#)