MSR142

Machine Learning to Predict Non-Compliance and Program Dropout in Patients Treated for Chronic Diseases: Unsupervised and Supervised Analyses from a Large Multinational Drug Access Program Database

Authors: Etienne Audureau, Joel Ladner, Ben Davis, and Joseph Saba

axios
TRANSFORMING ACCESS TO HEALTHCARE

OBJECTIVES

Patient non-compliance with treatment remains a global challenge, contributing to increased morbidity, mortality, and healthcare costs. Limited insight into underlying causes hampers targeted intervention design. This study applies machine learning to a large multi-country dataset to identify predictors of treatment non-compliance and program dropout in patients with chronic diseases.

METHODS

- We analyzed data from 29,959 patients enrolled between 2016 and 2024 in pharmaceutical access programs across 16 countries, covering drugs in 11 therapeutic areas.
- Descriptive and unsupervised clustering analyses were performed to evaluate correlations between individual- (demographics, clinical characteristics), program- (kind and extent of interactions between patients and program stakeholders), and country-level (GDP per capita) characteristics, and their associations with compliance and program dropout rates.
- Multivariable linear regression and logistic regression were used to assess the predictors of patient compliance (%) and dropout. Continuous covariates were modeled as linear predictors or using natural cubic splines, with the optimal functional form (linear to 5 knots) selected based on the Akaike Information Criterion (AIC) and likelihood ratio tests.
- Complementary analyses were performed using a random forest algorithm (R package *ranger*) to explore variable importance and potential non-linear interactions. Results were illustrated using variable importance plots and SHAP (SHapley Additive exPlanations) value visualizations.

RESULTS (Compliance)

Preliminary analysis found that the main disease areas were oncology (43.3%) and immunology (22.1%). Mean age was 48.9 years (±standard deviation 19.1; sex ratio M:F=0.92). The overall mean compliance was 46.2%±37.2.

Significant predictors of compliance included:

- Age: higher compliance in older and younger patients
- Treatment domain: highest in Endocrinology, Pulmonology, and Immunology
- Insurance status: highest compliance in insured patients
- Nb of treatment plans already implemented
- GDP per capita: higher compliance with increasing GDP per capita up to $10,000, then decreasing in the highest GDP per capita values, corresponding to the UAE and Saudi Arabia, where access program patients are likely not representative of the country's wealth
- Interactions between Axios and the patient at the inception of the program, with:
  - Increasing compliance with nb of patient interest interactions
  - U-curves for interactions around enrolment with generally better compliance around zero interactions or highest interactions

TABLES & FIGURES

Characteristic	N = 29,959[1]
Gender, women	15,564 (52.0%)
Age	48.9 ± 19.1 49.0 [35.0;64.0]
Insurance status	
Is insured	10,458 (34.9%)
Is not insured	12,656 (42.2%)
Other/Unknown	6,845 (22.8%)
Treatment	
Nb of treatment plans	1.34 ± 0.85 1.00 [1.00;1.00]
Overall Reported Compliance (%)	46.2 ± 37.2 41.0 [10.0;83.0]
Treatment domain	
Endocrinology	477 (1.6%)
Cardiology	1,774 (5.9%)
Dermatology	768 (2.6%)
Gastroenterology	1,279 (4.3%)
General	2,239 (7.5%)
Hematology	131 (0.4%)
Immunology	6,606 (22.1%)
Neurology	1,476 (4.9%)
Oncology	12,963 (43.3%)
Ophthalmology	1,585 (5.3%)
Pulmonology	580 (1.9%)
Unknown	81 (0.3%)
Interactions initiated by the patient at the 'Patient interest' step	0.70 ± 0.49 1.00 [0.000;1.00]
Interactions initiated by the patient at enrolment	2.42 ± 3.33 1.00 [0.000;4.00]
Interactions initiated by Axios at enrolment	6.0 ± 4.84 5.0 [3.00;7.3]
Country	
Country name	
Bulgaria	944 (3.2%)
Brazil	99 (0.3%)
Egypt	3,100 (10.3%)
India	1,722 (5.7%)
Indonesia	133 (0.4%)
Kuwait	1,804 (6.0%)
Lebanon	1,380 (4.6%)
Malaysia	3,604 (12.0%)
Mexico	489 (1.6%)
Morocco	539 (1.8%)
Philippines	1,177 (3.9%)
Saudi Arabia	1,297 (4.3%)
Thailand	2,853 (9.5%)
Ukraine	870 (2.9%)
United Arab Emirates	9,113 (30.4%)
Viet Nam	835 (2.8%)
GDP per capita (US $)	20129 ± 17244 11228 [3689;43982]

[1]n (%); Mean ± SD
Median [Q1;Q3]

Table 1. Descriptive statistics of the study population

Characteristic	Beta	95% CI	p-value
Age, years	*Spline, df=3; See Fig 1.*		<0.001
Gender, women	**0.8**	-0.02, 1.6	0.055
Treatment domain			
Endocrinology	**Ref.**	—	
Cardiology	**-12**	-16, -8.0	<0.001
Dermatology	**-9.5**	-14, -5.2	<0.001
Gastroenterology	**-7.9**	-12, -4.1	<0.001
General	**-11**	-15, -7.3	<0.001
Hematology	**-27**	-34, -20	<0.001
Immunology	**-7.5**	-11, -4.1	<0.001
Neurology	**-24**	-28, -20	<0.001
Oncology	**-13**	-17, -10	<0.001
Ophthalmology	**-13**	-17, -8.9	<0.001
Pulmonology	**-6.4**	-11, -2.0	0.005
Unknown	**-36**	-45, -28	<0.001
Insurance status			
Is insured	**Ref.**	—	
Is not insured	**-2.7**	-3.7, -1.6	<0.001
Other/Unknown	**-13**	-14, -12	<0.001
Nb of treatment plans	**4**	3.5, 4.5	<0.001
GDP per capita, US $	*Spline, df=4; See Fig 1.*		<0.001
Interactions initiated by Axios at enrolment	*Spline, df=3; See Fig 1.*		<0.001
Interactions initiated by the patient at the 'Patient interest' step	**7.4**	5.8, 9.1	<0.001
Interactions initiated by the patient at enrolment	*Spline, df=3; See Fig 1.*		<0.001

Abbreviation: CI = Confidence Interval; df = degrees of freedom.

Table 2. Predictors of patient compliance: results from multivariable linear regression modeling

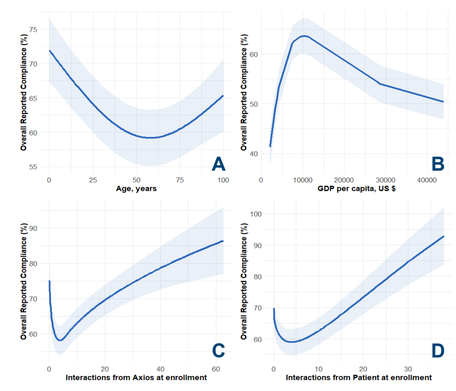

Figure 1. Nonlinear associations between compliance and covariates after spline transformation in linear regression model
A. Age, years;
B. GDP per capita, US $;
C. Interactions initiated by Axios at enrolment;
D. Interactions initiated by patients at enrolment.

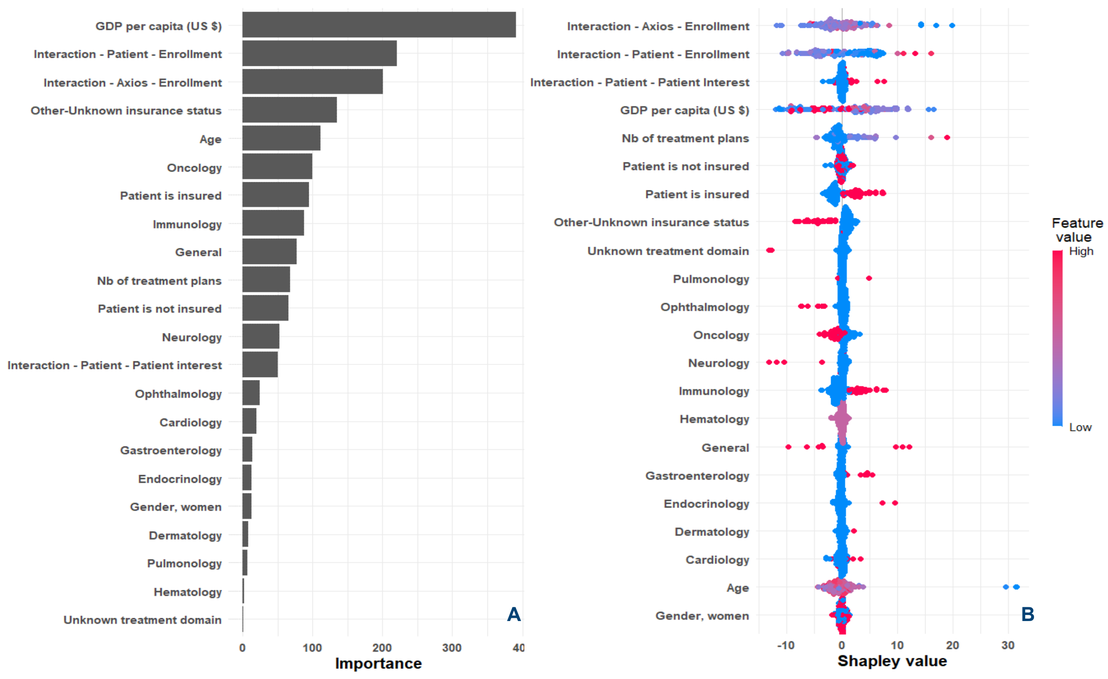

Figure 2.
Predictors of patient compliance by machine learning Random Forest algorithm: A. Variable importance; B. Shapley Values

Interpretation of the plot:
- X-axis (SHAP Value): Values to the right of zero indicate a positive contribution to the predicted compliance score (higher compliance). Values to the left indicate a negative contribution (lower compliance).
- Color Scale (Feature Value): The color of each point represents the original value of that feature for that observation.
- *Example: If high (red) values of predictor_A are mostly on the positive side of the SHAP axis, it means higher values of predictor_A lead to higher predicted compliance.*

CONCLUSION

These findings underscore the value of machine learning in identifying factors contributing to non-compliance and program dropout. The insights generated can support the design of tailored interventions to improve treatment compliance and retention across diverse patient populations.

Learn more about Axios International