A Machine Learning-Driven Multidimensional Evaluation of Tandospirone Formulations Based on Real-World Evidence

RWD855

ZOULinke[1,2] LIU Xinyu[2] WU Xingwei[2] HuMing[1]

1. West China School of Pharmacy, Sichuan University

2.Department of Pharmacy, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu

Background and Objective

- **BACKGROUND：** With the acceleration of healthcare informatization, electronic health records (EHRs) have accumulated vast amounts of data, providing abundant resources for real-world data (RWD) studies. However, the complex data structures of EHRs, coupled with inefficient and error-prone extraction of textual information such as progress notes, and issues like missing outcome data leading to bias, pose significant challenges.
- In recent years, machine learning and natural language processing (NLP) technologies have demonstrated significant potential in areas such as unstructured data extraction and missing data imputation.
- **OBJECTIVES:** To compare the effectiveness and safety of different formulations of Tandospirone, enabling precision therapeutic decision-making in patient subgroups.

Methods

● **Data Source & Extraction:**

Source: Outpatient and inpatient EHRs from patients diagnosed with anxiety disorders at a Grade A Tertiary Hospital in Sichuan Province, China (December 2018 - July 2023).

Inclusion Criteria: Patients with ≥2 consecutive clinical encounters; presence of diagnostic records, laboratory results, and medication information.

Exclusion Criteria: Patients lacking essential diagnostic, laboratory, or medication data.

● **Extraction Methods:**

Structured Data: Directly extracted from EHR databases.

Unstructured Data: Key variables captured using regular expression (regex) techniques for standardized retrieval.

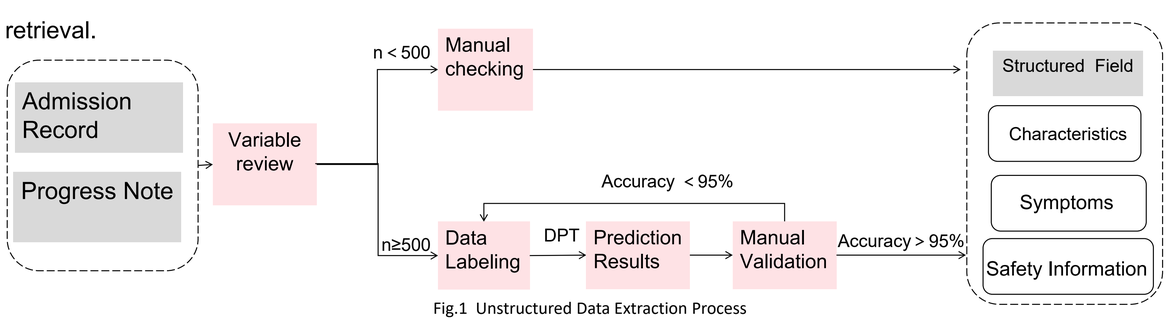

Fig.1 Unstructured Data Extraction Process

● **Anxiety Severity Assessment & Data Processing:**

① **Severity Determination:** Anxiety severity (Mild/Moderate/Severe) was classified using GAD-7 and SAS scales, integrated with clinical record review.

② **Outcome Transformation:** A One-versus-the-Rest (OVR) strategy converted the 3-level severity into three binary variables (Case vs. Non-Case for each level).

③ **Handling Missing Anxiety Scores**: After LASSO/BORUTA feature selection, missing data were imputed via KNN, RF, or fixed values; the dataset was then balanced with SMOTE/BSMOTE and split 8:2 into training/test sets. Eight machine learning were constructed to predict missing anxiety severity.

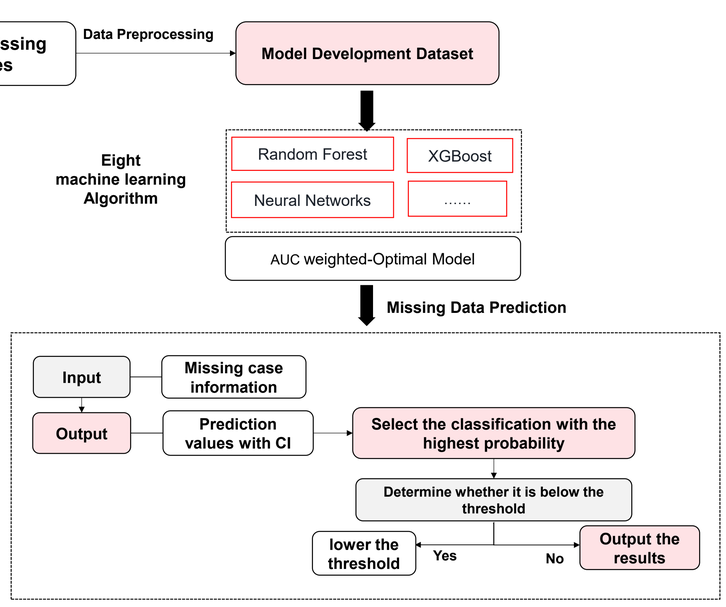

Fig.2 Flowchart for Predicting Cases with Missing Outcome Indicators

● **Effectiveness evaluation:** Efficacy was calculated as the proportion of cases with improved scale scores after 60±20 days of treatment (improved cases/assessable population).

● **Safety Evaluation:** "Unstructured Data Extraction" strategy was employed to structurally extract Adverse Drug Reactions (ADR) from clinical progress notes. ADR events occurring within 60 days of Tandospirone treatment with causality assessed as "possible or higher" were included.

● **Subgroup Analysis:** Treatment response rates were calculated based on the proportion of anxiety severity improvement, combined with multivariate analysis to identify influencing factors and conduct subgroup comparisons. Overall and subgroup ADR incidence rates were calculated, influencing factors were analyzed, and relative risk differences between formulations were compared.

Tab.1 Comparison of Safety/Efficacy in a Subgroup for Medication

Group	key factors: sex	age	BMI	...	...	Sample size	percentage	Safety/effectiveness indicators	Rank
1									1
b									2
...									...
n									n

Results

Results： This study analyzed 12,265 inpatient and 144,483 outpatient cases, achieving 95% data processing accuracy with an anxiety prediction model AUC >0.9. Tandospirone capsules and tablets showed comparable efficacy rates at 60±20 days (93.44% vs 92.87%). In efficacy evaluations, capsules outperformed tablets in 70.94% of subgroups (vs 7.30% for tablets). Safety analyses revealed no significant difference in ADR rates between formulations (1.53% vs 1.63%, p=1), demonstrating superiority in 82.69% of subgroups (363/439) the incidence of ADR with capsules compared to tablets (3.87%, 17/439).

Tab.2 Effective Rate and The incidence of ADR for Treatment with Tandospirone Capsules and Tablets

Evaluation dimension	Dosage Form	Sample Size (cases)	Efficacy Rate /Incidence of Adverse Reactions (%)	Chi-square Value	p-value
Effectiveness	Tablet	61051	93.44	0.1357	0.7126
	Capsule	421	92.87		
Safety	Tablet	491	1.63	0	1
	Capsule	7733	1.53		

Tab.3 Examples of relative safety of Tandolulone capsules and tablets within the same subgroup

Number	Age Group	Gender	Digestive System Diseases	Length of Hospital Stay	Ethnicity	Epithelial Cell Count	Formulation	ADR Incidence Rate (%)	Sample Size (cases)
1	18 - 40	Female	No	> 15 days	Yes	> 5	Capsule	16.67	22
2	18 - 40	Female	No	> 15 days	Yes	> 5	Tablet	9.09	11

Conclusion

Conclusion: This study revealed significant variations in the efficacy and safety profiles of Tandospirone capsules versus tablets across patient subgroups with distinct clinical characteristics, while demonstrating machine learning's utility in data extraction and missing value imputation. This finding underscore the imperative of personalized treatment strategies in clinical practice, and established a methodological foundation for health outcome evaluation.

*Email: Correspondence: huming@scu.edu.cn; First author: zlk1028@qq.com

Presented at ISPOR 2025, Sept. 28-30, 2025, Tokyo, Japan.