Development and Validation of a Risk Prediction Model of linezolid-induced thrombocytopenia in Vietnamese patients

Nhi Nguyen Ha An Tang Quoc

Friday, March 22, 2024

Write abstract here, note the indentation

1 Checklist

Table 1: TRIPOD-Cluster checklist of items to include when reporting a study developing or validating a multivariable prediction model using clustered data

Section/topic	Item No	Description	Draft date
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted	
Abstract	2	Provide a summary of research objectives, setting, participants, data source, sample size, predictors, outcome, statistical analysis, results, and conclusions*	

Section/topic	Item No	Description	Draft date
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the prediction model, including references to existing models, and the advantages of the study design*	
	3b	Specify the objectives, including whether the study describes the development or validation of the model*	
Methods			
Participants and data	4a	Describe eligibility criteria for participants and datasets*	
	4b	Describe the origin of the data, and how the data were identified, requested, and collected	
Sample size	5	Explain how the sample size was arrived at*	Mar 21
Outcomes and predictors	6a	Define the outcome that is predicted by the model, including how and when assessed*	Mar 21

Section/topic	Item No	Description	Draft date
	6b	Define all predictors used in developing or validating the model, including how and when measured*	
Data preparation	7a	Describe how the data were prepared for analysis, including any cleaning, harmonisation, linkage, and quality checks	
	7b	Describe the method for assessing risk of bias and applicability in the individual clusters (eg, using PROBAST)	
	7c	For validation, identify any differences in definition and measurement from the development data (eg, setting, eligibility criteria, outcome, predictors)*	
	7d	Describe how missing data were handled*	
Data analysis	8a	Describe how predictors were handled in the analyses	
	8b	Specify the type of model, all model building procedures (eg, any predictor selection and penalisation), and method for validation*	

Section/topic	Item No	Description	Draft date
	8c	Describe how any heterogeneity across clusters (eg, studies or settings) in model parameter values was handled	
	8d	For validation, describe how the predictions were calculated	
	8e	Specify all measures used to assess model performance (eg, calibration, discrimination, and decision curve analysis) and, if relevant, to compare multiple models	
	8f	Describe how any heterogeneity across clusters (eg, studies or settings) in model performance was handled and quantified	
	8g	Describe any model updating (eg, recalibration) arising from the validation, either overall or for particular populations or settings*	
Sensitivity analysis	9	Describe any planned subgroup or sensitivity analysis—eg, assessing performance according to sources of bias, participant characteristics, setting	

Section/topic	Item No	Description	Draft date
Results			
Participants and datasets	10a	Describe the number of clusters and participants from data identified through to data analysed; a flowchart might be helpful*	
	10b	Report the characteristics overall and where applicable for each data source or setting, including the key dates, predictors, treatments received, sample size, number of outcome events, follow-up time, and amount of missing data*	
	10c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors, and outcome)	
Risk of bias	11	Report the results of the risk-of-bias assessment in the individual clusters	

Section/topic	Item No	Description	Draft date
Model development and specification	12a	Report the results of any assessments of heterogeneity across clusters that led to subsequent actions during the model's development (eg, inclusion or exclusion of particular predictors or clusters)	
	12b	Present the final prediction model (ie, all regression coefficients, and model intercept or baseline estimate of the outcome at a given time point) and explain how to use it for predictions in new individuals*	
Model performance	13a	Report performance measures (with uncertainty intervals) for the prediction model, overall and for each cluster	
	13b	Report results of any heterogeneity across clusters in model performance	
Model updating	14	Report the results from any model updating (including the updated model equation and subsequent performance), overall and for each cluster*	

Section/topic	Item No	Description	Draft date
Sensitivity analysis	15	Report results from any subgroup or sensitivity analysis	
Discussion			
Interpretation	16a	Give an overall interpretation of the main results, including heterogeneity across clusters in model performance, in the context of the objectives and previous studies*	
	16b	For validation, discuss the results with reference to the model performance in the development data, and in any previous validations	
	16c	Discuss the strengths of the study and any limitations (eg, missing or incomplete data, non-representativeness, data harmonisation problems)	
Implications	17	Discuss the potential use of the model and implications for future research, with specific view to generalisability and applicability of the model across different settings or (sub)populations	
Other information			

Section/topic	Item No	Description	Draft date
Supplementary information	18	Provide information about the availability of supplementary resources (eg, study protocol, analysis code, datasets)*	
Funding	19	Give the source of funding and the role of the funders for the present study	

2 Introduction

2.1 Background and objectives

Linezolid is an oxazolidinones antibiotics [1].

3 Methods

3.1 Participants and data

3.2 Sample size

Previous studies developing logistic regression models for LI-TP risk predictions have included 4-6 predictors in their final models [2]; [3]; [4]; [5]. We expect to include about as many candidate predictors, based on results from the expert opinion survey and the Bayesian Model Selection algorithm see 3.3. Some of the candidate predictors might be continuous, which may potentially require non-linear modelling and therefore slightly increase the number of variables.

A general rule of thumb is for at least 10 events be available for each candidate predictor considered in a prediction model [6]. We have a total of 816 eligible patients and 264 of those have experienced the outcome. If the number of candidate predictors is 7, we would have 37 events per candidate predictor, which is considerably greater than the minimum number required. Even if the number of parameters screened is 20, we would still have 13 events per candidate predictor.

However, the aforementioned rule of thumb have generated some debate in the literature, with recent results suggesting that event per variable criterion is too simplistic and has no strong

relation to the predictive performance of a model. Riley et al [7] proposed a different set of criteria to estimate minimum sample size for models developed using logistic regression, which are the following:

1. Small optimism in predictor effect estimates, defined as a global shrinkage factor of >= 0.9.
2. Small absolute difference of <= 0.05 in the model's apparent and adjusted Nagelkerke's R-squared.
3. Precise estimation of the overall risk in the population.

Criteria 1 and 2 aims to reduce the potential of overfitting. Criteria 3 aims to ensure the overall risk is estimated precisely.

3.2.1 Step 1: Choose the number of candidate predictors of interest for inclusion in the model, and calculate the corresponding number of predictor parameters (p)

Note that one predictor may require two or more parameters. For example, a k-category predictor requires k-1 parameters and a continuous predictor model with a non-linear trend requires more than one parameter to be estimated. Also include any potential interaction terms towards the total p.

When using a predictor selection method, p should be defined as the total number of parameters screened, and not just the subset that are included in the final model.

Assuming maximum total p to be 20.

> **Note**
>
> The value of p is assumed to be no larger than 20 because univariate regression shows there are 20 variables that are significantly correlated with the outcome.

Source: Article Notebook

3.2.2 Step 2: Choose sensible values for $R^2_{CS\_adj}$ and max($R^2_{CS\_app}$) based on previous studies where R^2_{CS} is the Cox-Snell R^2 statistic.

The value of max($R^2_{CS\_app}$) is based on the overall prevalence or overall rate of the outcome in the population of interest. The incidence of LI-TP in patients treated with linezolid was estimated to be 37% in a meta-analysis by Zhao et al [8].

The value of $R^2_{CS\_adj}$ could be based on that for a previously published model in the same setting and population (with similar outcome definition). However, as previous studies does not provide any information to identify a sensible value of the minimum expected Cox-Snell R^2, the value $R^2_{CS\_adj}$ will be assumed to correspond to a $R^2_{Nagelkerke}$ of 0.15.

```
[1] 0.2048324
```

Source: Article Notebook

3.2.3 Step 3: Criterion 1

Calculate the sample size required to ensure Van Houwelingen's global shrinkage factor (S_{VH}) is close to 1. A value of S_{VH} >= 0.90 is generally recommended, which reflects a small amount of overfitting during model development.

```
[1] 775
```

Source: Article Notebook

We see that 775 participants are required to meet criterion 1.

3.2.4 Step 4: Criterion 2

Calculate the shrinkage factor (S_{VH}) required to ensure a small absolute difference of <= 0.05 in the developed model's apparent and adjusted Nagelkerke's R^2. Then derive the required sample size conditional on this value of S_{VH}.

```
[1] 478
```

Source: Article Notebook

We see that 478 participants are required to meet criterion 2.

3.2.5 Step 5: Criterion 3

Calculate the sample size required to ensure a precise estimate of the overall risk in the population. The suggested absolute margin of error is <= 0.05.

```
[1] 359
```

Source: Article Notebook

We see that 359 participants are required to meet criterion 3.

3.2.6 Step 6: Final sample size

The required minimum sample size is the maximum value from steps 3 to 5, to ensure that each of criteria 1 to 3 are met.

```
[1] 775
```

Source: Article Notebook

The final estimate of minimum sample size is 775, therefore our data is sufficient for model development with at most 20 parameters.

3.3 Outcomes and predictors

3.3.1 6a. Define the outcome that is predicted by the model, including how and when assessed

The outcome of interest is linezolid-induced thrombocytopenia, defined as (i) a platelet count of < 112.5 x 10^9 cells/L (75% of the lower limit of normal) for patients with a baseline platelet count in the normal range; (ii) A reduction in platelet count of 25% from the baseline value for patients with a baseline platelet count of < 150 x 10^9 cells/L [9]; [5]; [10].

Normal platelet count is defined as 150-450 x 10^9 cells/L. Baseline platelet count is defined as the last recorded PLT value before the start of linezolid therapy. Participants are considered to have met the outcome if they have a platelet count that meets the above criteria at any time during linezolid therapy or within 5 days after the end of therapy.

> **Warning**
>
> Thrombocytopenia may occur within a few days after stopping LZD, when the drug hasn't been completely eliminated. However, it is unknown exactly how long after stopping LZD can a TP event still be attributed to LZD use. The value of 5 days has been chosen *almost* arbitrarily.
> Our rationale is that after 5 days (120 hrs), LZD is guaranteed to be completely eliminated in all patients, as the longest $t_{1/2}$ is 8.3 ± 2.4 hrs in end-stage renal disease patients, +3 SD would be ~16 hrs, so 120 hrs is >7 half-lives, therefore in patients with the worst clearance, 99% of them would have 99% of the drug eliminated from their systems. Furthermore, trough LZD concentration (C_{min}) has previously been identified as a predictor of LI-TP development, and LI-TP itself is mostly reversible after discontinuation, so we would argue that any TP events that occur after LZD has been eliminated from the system would not be related to LZD use.

3.3.2 6b. Define all predictors used in developing or validating the model, including how and when measured

3.4 Data preparation

3.4.1 7a. Describe how the data were prepared for analysis, including any cleaning, harmonisation, linkage, and quality checks

3.5 Data analysis

3.6 Sensitivity analysis

4 Results

4.1 Participants and datasets

4.2 Risk of bias

4.3 Model development and specification

4.4 Model performance

4.5 Model updating

4.6 Sensitivity analysis

5 Discussion

5.1 Interpretation

5.2 Implications

6 Other information

6.1 Supplementary information

6.2 Funding

6.3 References

1. Cazavet J, Bounes FV, Ruiz S, Seguin T, Crognier L, Rouget A, et al. Risk factor analysis for linezolid-associated thrombocytopenia in critically ill patients. Eur J Clin Microbiol Infect Dis

[Internet]. 2020;39:527–38. Available from: https://www.ncbi.nlm.nih.gov/pubmed/31853741

2. Liu Y, Liu T, Wei G, Yan P, Fang X, Xie L. Analysis of risk factors and establishment of risk prediction model for linezolid-associated thrombocytopenia. Medical Journal of Chinese People's Liberation Army [Internet]. 2021;46. Available from: https://d.wanfangdata.com.cn/periodical/jfjyxzz202108006

3. Duan L, Zhou Q, Feng Z, Zhu C, Cai Y, Wang S, et al. A Regression Model to Predict Linezolid Induced Thrombocytopenia in Neonatal Sepsis Patients: A Ten-Year Retrospective Cohort Study. Front Pharmacol [Internet]. 2022;13:710099. Available from: https://www.ncbi.nlm.nih.gov/pubmed/35185555

4. Qin Y, Chen Z, Gao S, Pan MK, Li YX, Lv ZQ, et al. Development and Validation of a Risk Prediction Model of Linezolid-induced Thrombocytopenia in Elderly Patients [Internet]. In Review; 2021 Jun. Available from: https://www.researchsquare.com/article/rs-582799/v1

5. Xu J, Lu J, Yuan Y, Duan L, Shi L, Chen F, et al. Establishment and validation of a risk prediction model incorporating concentrations of linezolid and its metabolite PNU142300 for linezolid-induced thrombocytopenia. The Journal of Antimicrobial Chemotherapy. 2023;78:1974–81.

6. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. Journal of Clinical Epidemiology [Internet]. 1995;48:1503–10. Available from: https://www.jclinepi.com/article/0895-4356(95)00048-8/abstract

7. Riley RD, Snell KI, Ensor J, Burke DL, Harrell Jr FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Statistics in Medicine [Internet]. 2019;38:1276–96. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.7992

8. Zhao X, Peng Q, Hu D, Li W, Ji Q, Dong Q, et al. Prediction of risk factors for linezolid-induced thrombocytopenia based on neural network model. Frontiers in Pharmacology [Internet]. 2024 [cited 2024 Feb 27];15. Available from: https://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2024.1292828

9. Zyvox prescribing information [Internet]. Available from: https://labeling.pfizer.com/showlabeling.aspx?id=649

10. Kawasuji H, Tsuji Y, Ogami C, Kimoto K, Ueno A, Miyajima Y, et al. Proposal of initial and maintenance dosing regimens with linezolid for renal impairment patients. BMC Pharmacology and Toxicology [Internet]. 2021 [cited 2024 Feb 26];22:13. Available from: https://doi.org/10.1186/s40360-021-00479-w