



University of Glasgow | School of
Computing Science

THE AWARDS
2020

UNIVERSITY
OF THE YEAR

A Guide to Model Validation in Clinical Decision Support Systems

Dr. Fani Deligianni,

fani.deligianni@glasgow.ac.uk

Lecturer (Assistant Professor)

Lead of the Computing Technologies for Healthcare Theme

<https://www.gla.ac.uk/schools/computing/staff/fanideligianni>

WORLD
CHANGING
GLASGOW



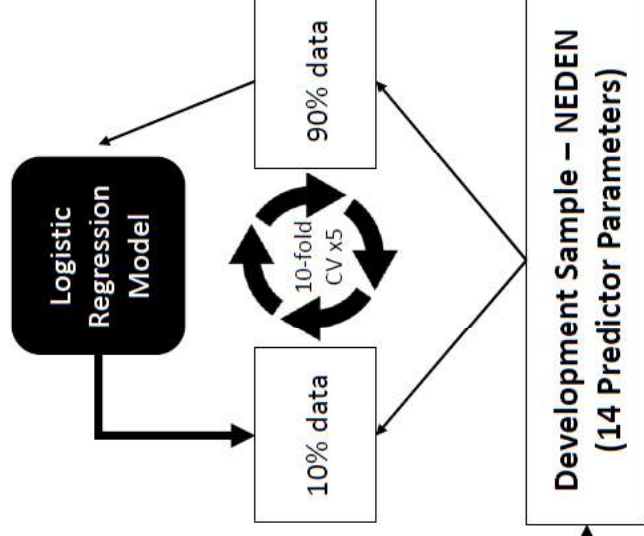
ABCD Guide

- A. Calibration-in-the-large, or the model intercept
- B. Calibration slope
- C. Discrimination with the Receiver Operating Characteristic curve
- D. Clinical usefulness with decision-curve analysis



Example in First Episode Psychosis

Internal Validation by nested CV



Pre-Processing Steps

- Select candidate predictors based on previous literature and expert knowledge.
- Multiply impute development and validation datasets ($m=10$).
- Standardise predictors.

External Validation

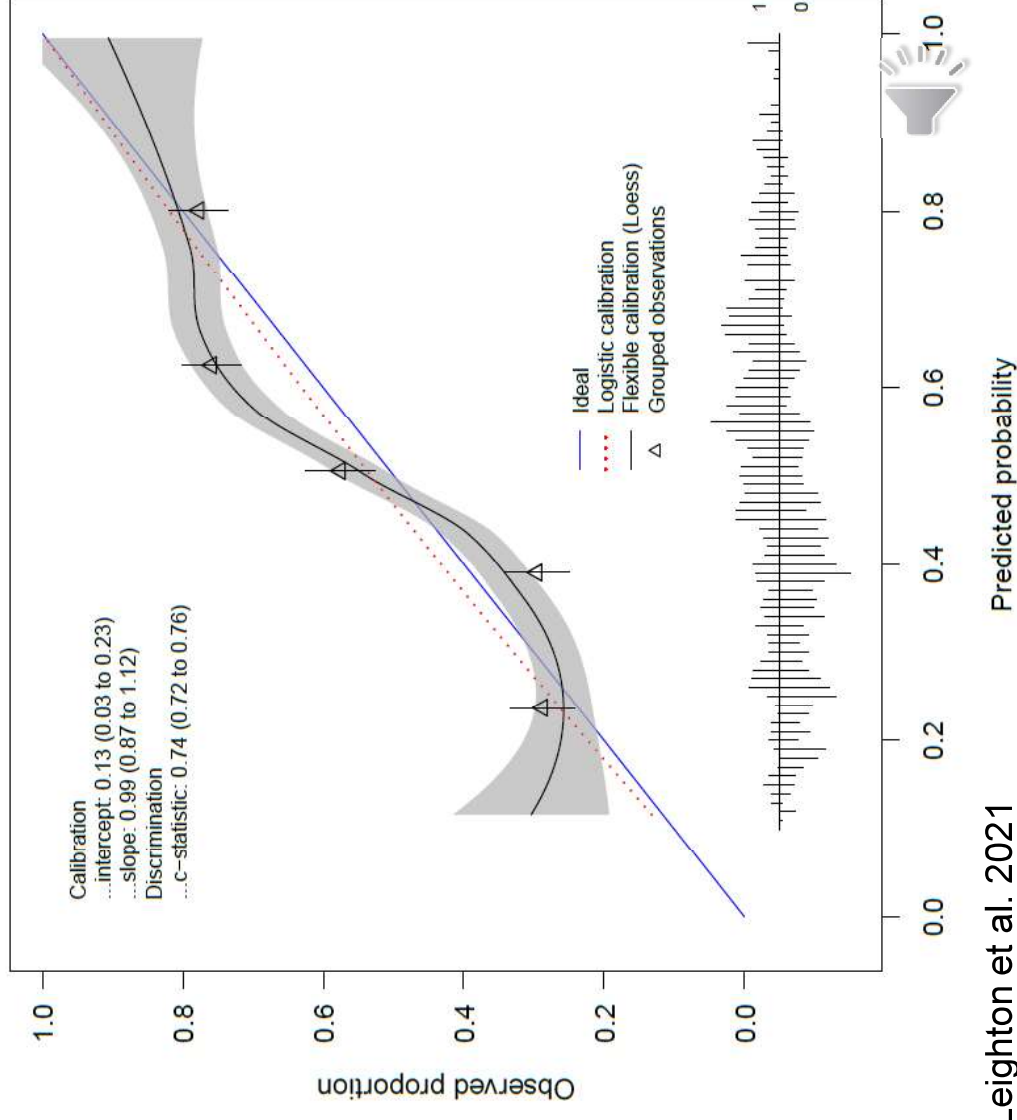
Validation Sample – Outlook

Leighton et al. Development and validation of a non-remission risk prediction model in First Episode Psychosis: An analysis of two longitudinal studies, 2021

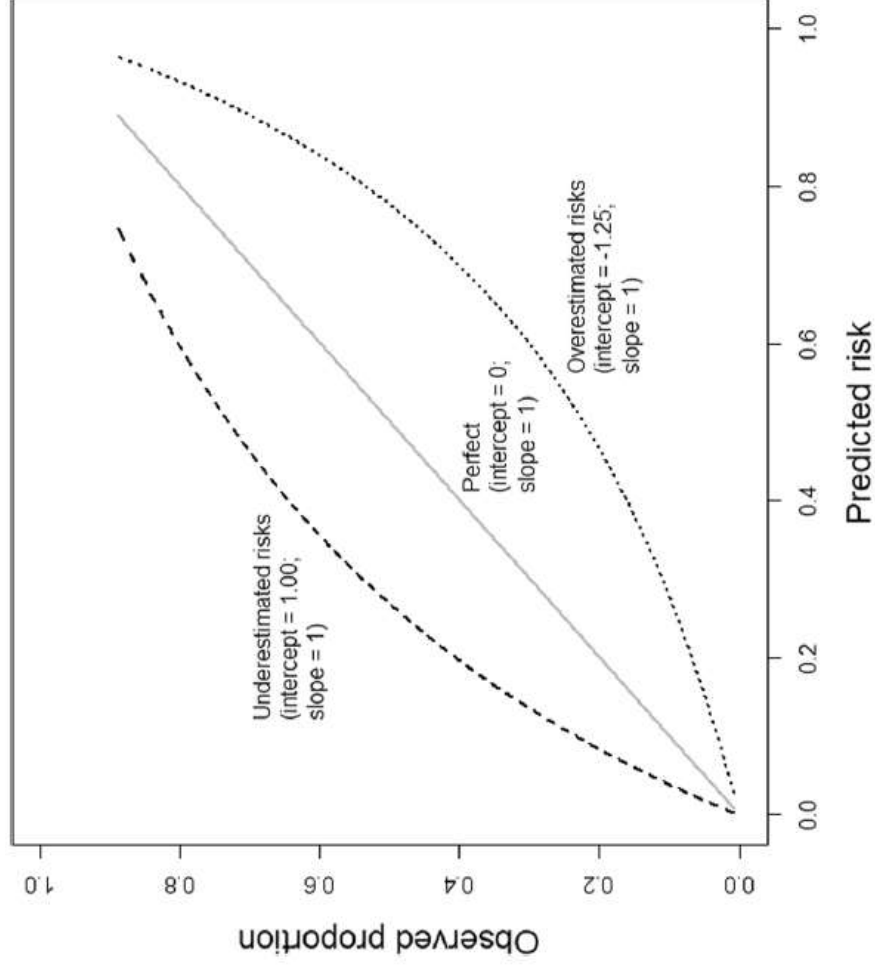


Example Calibration - Discrimination

- Calibration refers to the agreement between observed outcomes and predictions
- Calibration-in-the-large – external validation
- Calibration slope – internal validation



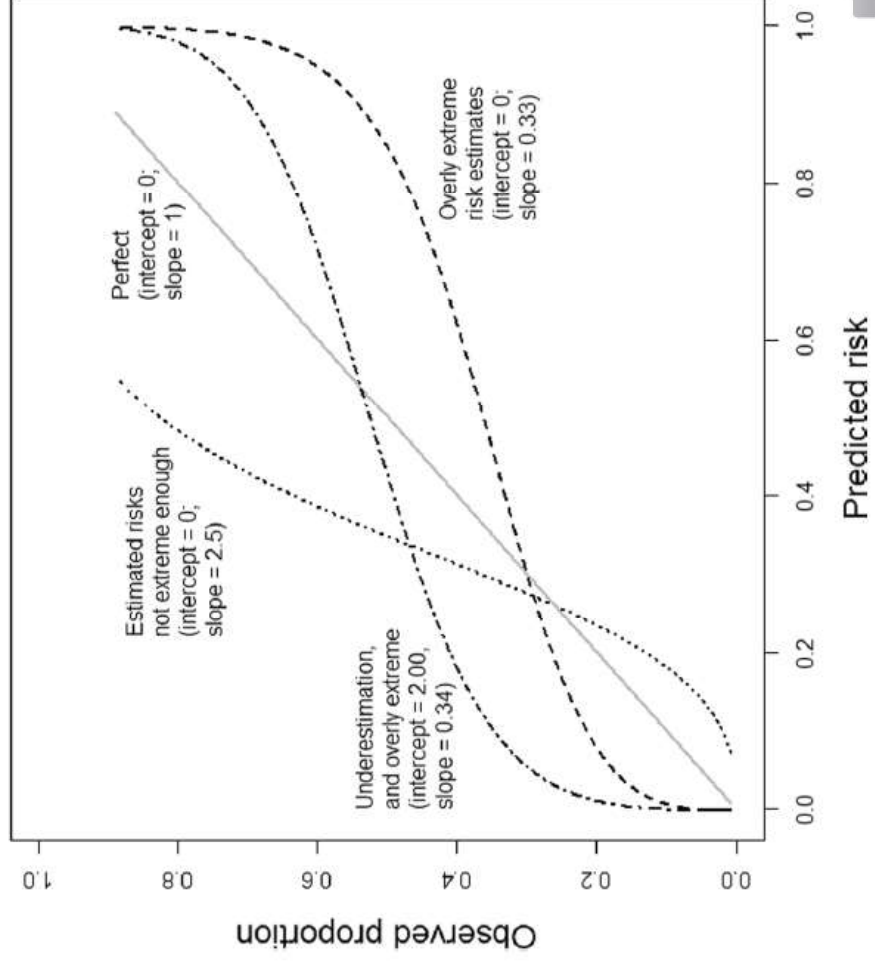
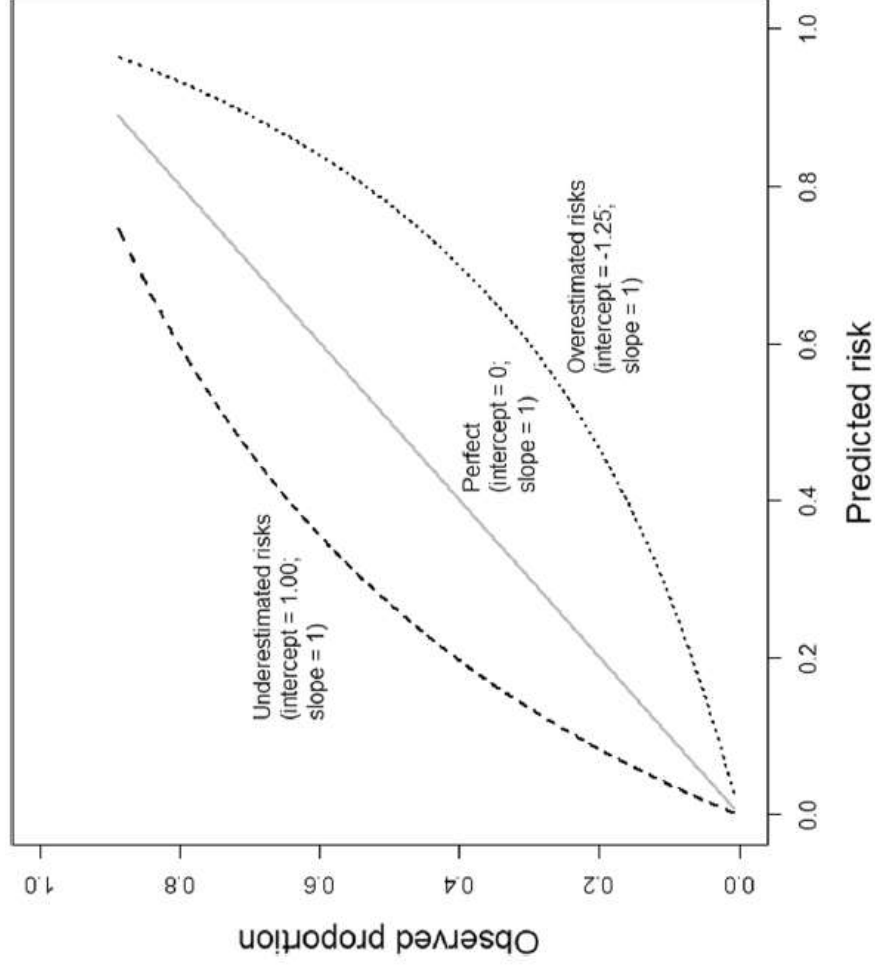
Assessing Calibration



Calster et al. Calibration: the Achilles heel of predictive analytics, 2019



Examples of Extreme Calibration



Calster et al. Calibration: the Achilles heel of predictive analytics, 2019



Clinical Consequences

- Overestimation of risks/probabilities affects the decision of patients
- Poor calibration reduces clinical usefulness
- Uncalibrated model are not accepted by medical professionals



Reasons of Poor Calibration

- Differences in patients population from one centre to another
- Patients' populations change over time
- Related to algorithm – overfitting
- Measurement errors in clinical data
- Intra and inter-observer variability

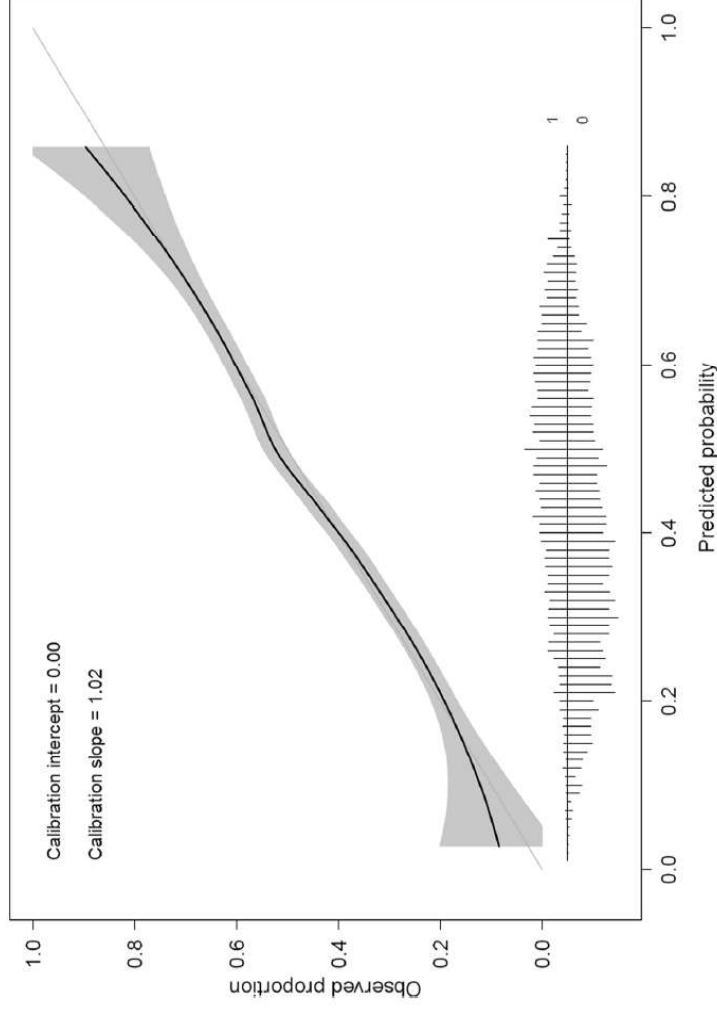
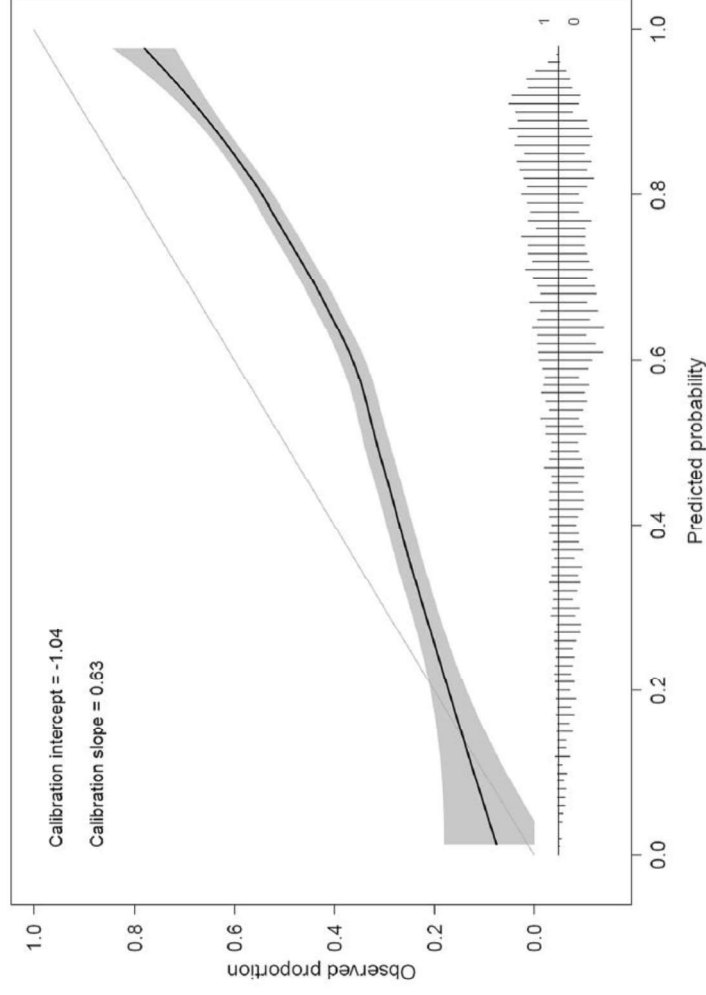


Correct Poor Calibration

- Avoid overfitting with simpler models and penalization **but**
- It does not improve uncertainty and it is not a solution for everything
- Update the regression coefficients
 - Change the intercept to correct calibration-in-the-large
 - Refitting the algorithm
 - Continuous updating



Correct Poor Calibration



Calster et al. Calibration: the Achilles heel of predictive analytics, 2019



Summary

- Calibration, Discrimination and Clinical Usefulness are of paramount importance in assessing clinical decision systems
- Calibration is important in clinical decision systems because they affect patients decision
- Calibration-in-the-large is important in external validation
- Several predictive models in AI do not examine calibration performance
- Performance is assessed via Receiver Operative Characteristic in order to take into consideration different thresholds.



References

- Steyerberg et al. 'Towards better clinical prediction models: seven steps for development and an ABCD for validation', European Heart Journal, 2014.
- Calster et al. Calibration : the Achilles heel of predictive analytics, BMC Medicine, 2019.
- Leighton et al. Development and validation of multivariable prediction models of remission, recovery, and quality of life outcomes in people with first episode psychosis: a machine learning approach, The Lancet: Digital Health, 2019.
- Leighton et al. Development and validation of a non-remission risk prediction model in First Episode Psychosis: An analysis of two longitudinal studies, BJPsych Open, 2021.