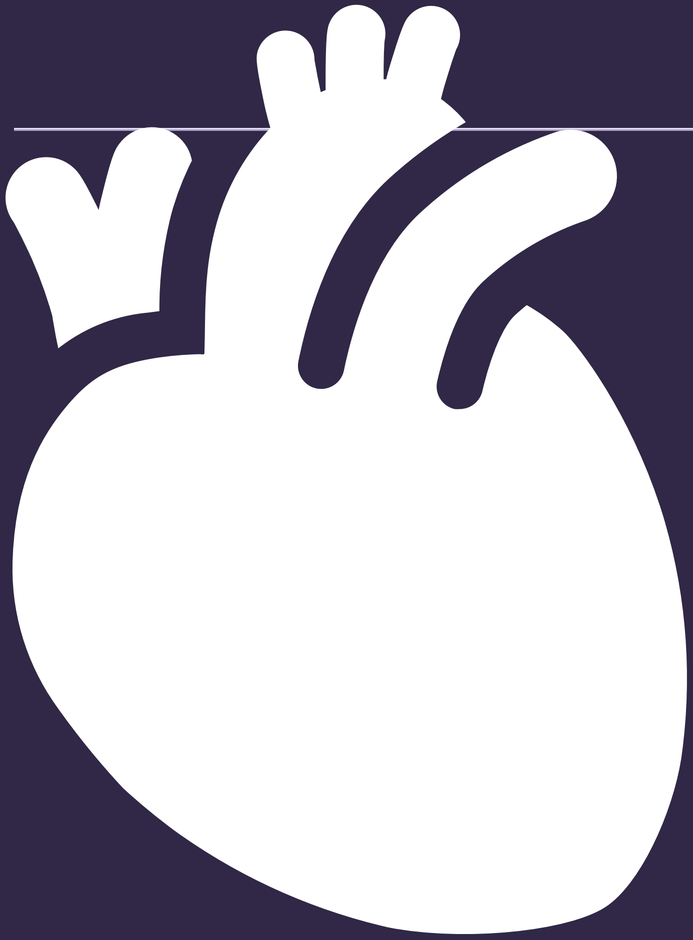
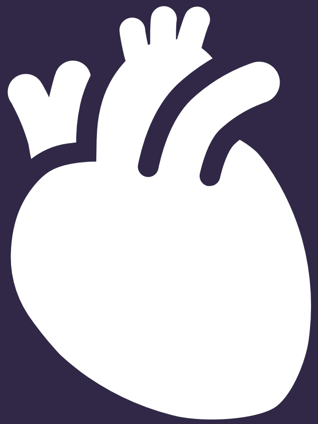
A suite of programs for clinical prediction modelling

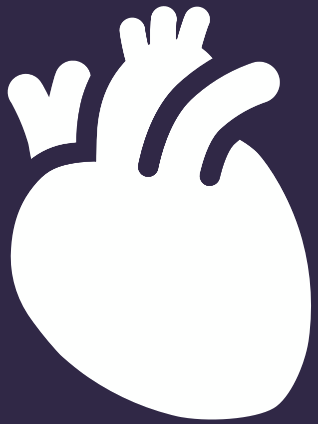


Dr Joie Ensor

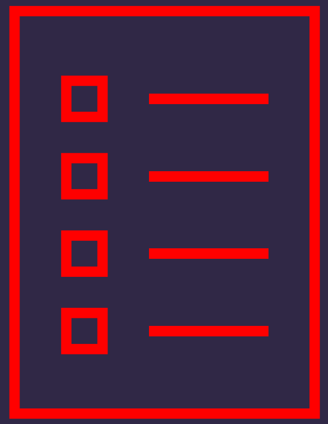
Email: j.ensor@bham.ac.uk
Twitter: @joie_ensor

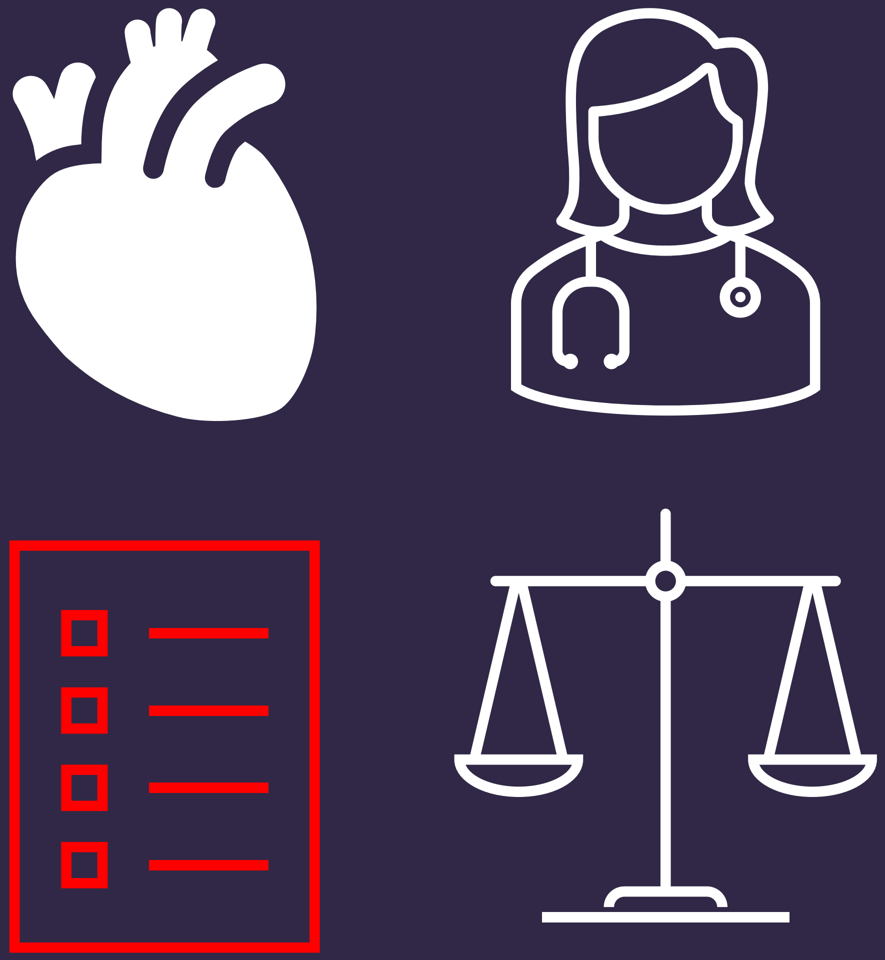
UNIVERSITY OF BIRMINGHAM

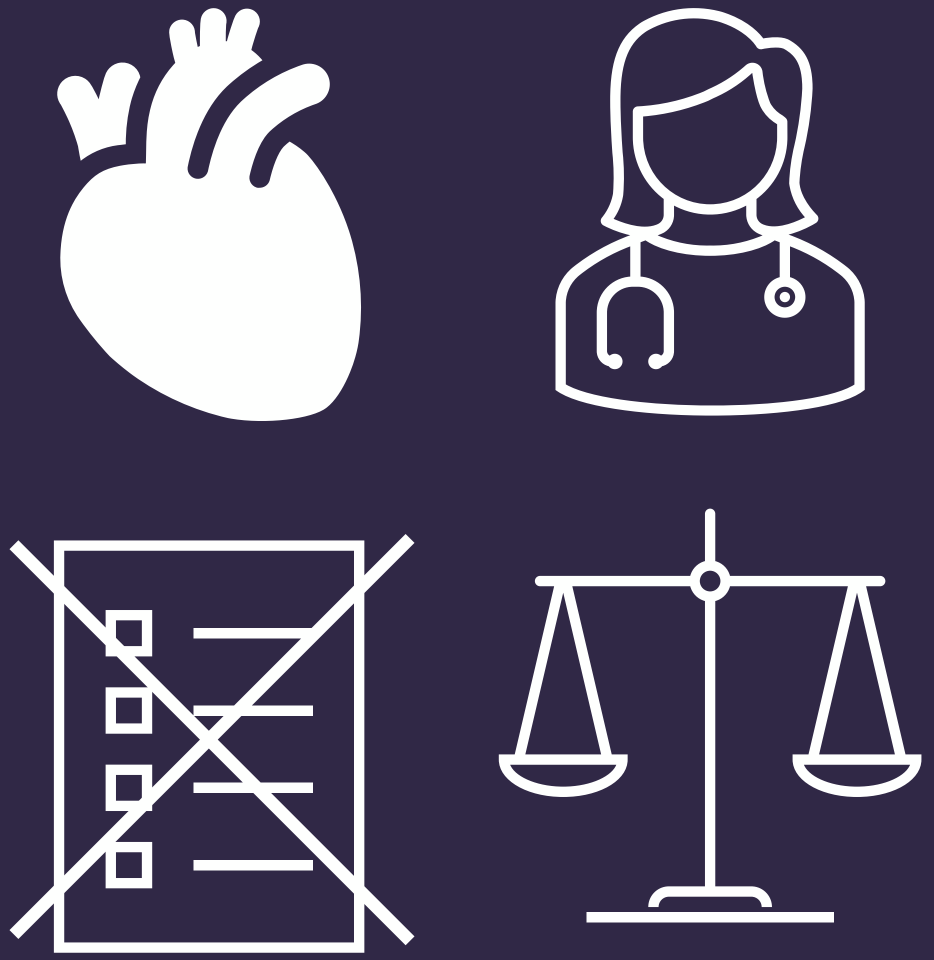


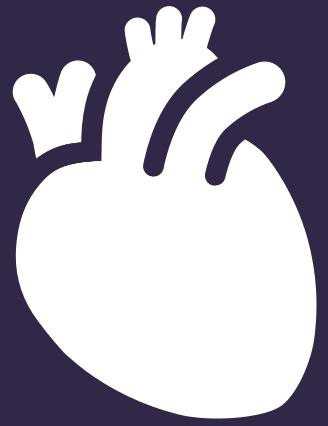




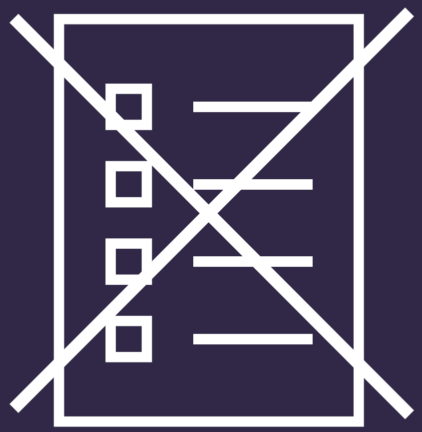













Accurate
&
Reliable

pm-suite

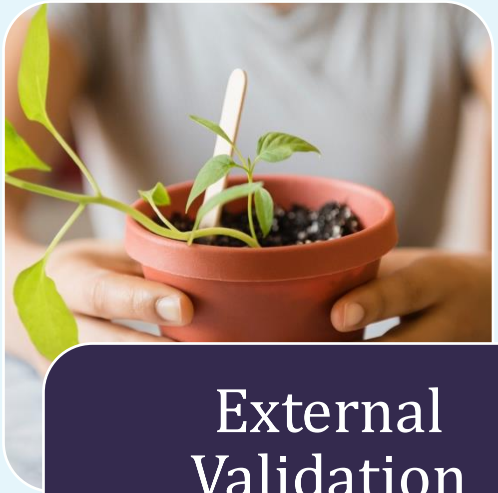
- Aid across the stages of clinical prediction modelling
- Methodology & TRIPOD embedded
- Useful for:
  - Design
  - Analysis
  - Reporting

CPM stages

Development

&

Internal Validation



External Validation

&

Updating

Implementation

&

Impact

Model *development*

- Variable selection
- Functional forms

Development

&

Internal Validation

Internal validation of our development process

- Assess model's validity within the same population
- Bootstrapping or cross-validation
- Quantify optimism
- Adjust our model

Development

&

Internal Validation

Measure the *performance* of the model

- At development performance estimates are optimistic
- After internal validation we adjust performance measures for optimism

Development

&

Internal Validation

External validation

- Assess model’s validity in patients separate from the first stage
- Models are developed to be applied in *new* individuals, so their value depends on their performance outside of the development sample

External Validation

&

Updating

Model performance

- How *accurate* & *reliable* is the model?
- Assess model *reproducibility* or *transportability*

External Validation

&

Updating

Model *updating*

- Inadequate performance could indicate updating

- Adjust the model to improve accuracy & reliability in a new setting/population

External Validation

&

Updating

pm-suite

Development
&
Internal Validation

External Validation
&
Updating

pmsampsize
pmintval

pmvalsampsize
pmupdate

pmstats
pmcalplot

pmiecv
pmcstat

pmsplot
pmmeta

“Only nine of 119 studies (8%) reported a sample size calculation”

Dhiman et al. 2023

Sample size considerations

- We want to have a large enough sample size to develop a model that predicts as accurately as we can

- Important when conducting a **prospective** study
  - How many individuals do I need to collect?

- Important when using **existing data**
  - Is my available data large enough?
  - How many predictors can I consider?

pmsampsize

- Minimum sample size required for developing a prediction model

- Calculates sample size that is needed to,
    - minimise potential overfitting
    - estimate parameters precisely (e.g., intercept)

- Implements a series of closed form solutions

```
. pmsampsize, type(b) prevalence(0.05) parameters(25) nagrsquared(0.15)
NB: Assuming 0.05 acceptable difference in apparent & adjusted R-squared
NB: Assuming 0.05 margin of error in estimation of intercept
NB: Events per Predictor Parameter (EPP) assumes prevalence = .05

             |  Samp_size  Shrinkage  Parameter     CS_Rsq    Max_Rsq    Nag_Rsq        EPP
-------------+-----------------------------------------------------------------------------
  Criteria 1 |       4466         .9         25       .049       .328        .15       8.93
  Criteria 2 |       1476       .749         25       .049       .328        .15       2.95
  Criteria 3 |         73          .         25       .049       .328        .15        .15
-------------+-----------------------------------------------------------------------------
    Final SS |       4466         .9         25       .049       .328        .15       8.93

Minimum sample size required for new model development based on user inputs = 4466,
with 224 events (assuming an outcome prevalence = .05), and an EPP = 8.93

Criteria 1 - small overfitting defined as expected shrinkage of predictor effects by 10% or less
Criteria 2 - small absolute difference in model's apparent and adjusted Nagelkerke's R-squared
Criteria 3 - precise estimation of the average outcome risk in the population
```

```
. pmsampsize, type(b) prevalence(0.05) parameters(25) nagrsquared(0.15)
NB: Assuming 0.05 acceptable difference in apparent & adjusted R-squared
NB: Assuming 0.05 margin of error in estimation of intercept
NB: Events per Predictor Parameter (EPP) assumes prevalence = .05

             |  Samp_size  Shrinkage  Parameter     CS_Rsq    Max_Rsq    Nag_Rsq        EPP
-------------+-----------------------------------------------------------------------------
  Criteria 1 |       4466         .9         25       .049       .328        .15       8.93
  Criteria 2 |       1476       .749         25       .049       .328        .15       2.95
  Criteria 3 |         73          .         25       .049       .328        .15        .15
-------------+-----------------------------------------------------------------------------
    Final SS |       4466         .9         25       .049       .328        .15       8.93

Minimum sample size required for new model development based on user inputs = 4466,
with 224 events (assuming an outcome prevalence = .05), and an EPP = 8.93

Criteria 1 - small overfitting defined as expected shrinkage of predictor effects by 10% or less
Criteria 2 - small absolute difference in model's apparent and adjusted Nagelkerke's R-squared
Criteria 3 - precise estimation of the average outcome risk in the population
```

```
. pmsampsize, type(b) prevalence(0.05) parameters(25) nagrsquared(0.15)
NB: Assuming 0.05 acceptable difference in apparent & adjusted R-squared
NB: Assuming 0.05 margin of error in estimation of intercept
NB: Events per Predictor Parameter (EPP) assumes prevalence = .05

             |  Samp_size  Shrinkage  Parameter     CS_Rsq    Max_Rsq    Nag_Rsq        EPP
-------------+-----------------------------------------------------------------------------
  Criteria 1 |       4466         .9         25       .049       .328        .15       8.93
  Criteria 2 |       1476       .749         25       .049       .328        .15       2.95
  Criteria 3 |         73          .         25       .049       .328        .15        .15
-------------+-----------------------------------------------------------------------------
    Final SS |       4466         .9         25       .049       .328        .15       8.93

Minimum sample size required for new model development based on user inputs = 4466,
with 224 events (assuming an outcome prevalence = .05), and an EPP = 8.93

Criteria 1 - small overfitting defined as expected shrinkage of predictor effects by 10% or less
Criteria 2 - small absolute difference in model's apparent and adjusted Nagelkerke's R-squared
Criteria 3 - precise estimation of the average outcome risk in the population
```

```
. pmsampsize, type(b) prevalence(0.05) parameters(25) nagrsquared(0.15)
NB: Assuming 0.05 acceptable difference in apparent & adjusted R-squared
NB: Assuming 0.05 margin of error in estimation of intercept
NB: Events per Predictor Parameter (EPP) assumes prevalence = .05

             |  Samp_size  Shrinkage  Parameter     CS_Rsq    Max_Rsq    Nag_Rsq        EPP
-------------+-----------------------------------------------------------------------------
  Criteria 1 |       4466         .9         25       .049       .328        .15       8.93
  Criteria 2 |       1476       .749         25       .049       .328        .15       2.95
  Criteria 3 |         73          .         25       .049       .328        .15        .15
-------------+-----------------------------------------------------------------------------
    Final SS |       4466         .9         25       .049       .328        .15       8.93

Minimum sample size required for new model development based on user inputs = 4466,
with 224 events (assuming an outcome prevalence = .05), and an EPP = 8.93

Criteria 1 - small overfitting defined as expected shrinkage of predictor effects by 10% or less
Criteria 2 - small absolute difference in model's apparent and adjusted Nagelkerke's R-squared
Criteria 3 - precise estimation of the average outcome risk in the population
```

- Options to specify anticipated performance of new model include:
  - Cox-Snell R-squared
  - Nagelkerke's R-squared
  - C statistic

```
. pmsampsize, type(b) prevalence(0.05) parameters(25) cstat(0.79)

Given C-statistic = .79 & prevalence = .05

Cox-Snell R-sq = 0.0586
```

Sample size considerations

- We want to have a large enough sample size to develop a model that predicts as accurately as we can

- Important when conducting a **prospective** study
  - How many individuals do I need to collect?

- Important when using **existing data**
  - Is my available data large enough?
  - How many predictors can I consider?

```
. pmsampsize, type(b) prevalence(0.05) rsquared(0.059) n(3688)
NB: Assuming 0.05 acceptable difference in apparent & adjusted R-squared
NB: Events per Predictor Parameter (EPP) assumes prevalence = .05

             |  Samp_size  Shrinkage  Parameter        Rsq    Max_Rsq    Nag_Rsq        EPP
-------------+-------------------------------------------------------------------------------
  Criteria 1 |       3688         .9         25       .059       .328        .18       7.38
  Criteria 2 |       3688       .783         62       .059       .328        .18       2.97
Criteria 3 * |       3688         .9         25       .059       .328        .18       7.38
-------------+-------------------------------------------------------------------------------
       Final |       3688         .9         25       .059       .328        .18       7.38

Maximum number of predictor parameters that could be estimated during new model development
based on user inputs = 25, with 185 events (assuming an outcome prevalence = .05) & an EPP = 7.38

* 95% CI for overall risk = (.043, .057), for true value of .05, sample size n=3688
Absolute margin of error = .007
```

“An explanation of sample size was reported in only 9% of validation studies”

Collins et al. 2014

Sample size for validation studies

What do we want?

We want to have a large enough sample size to …

Development

- *develop a model that predicts as accurately as we can*

Validation

- *accurately and precisely estimate model performance*

pmvalsampsize

- Minimum sample size required for external validation of a prediction model

- Calculates sample size needed to ensure precise estimation of key measures of prediction model performance

```
. pmvalsampsize, type(b) prevalence(0.05) cstat(0.74) lpnormal(-3.25, 0.9) graph
Normal LP distribution with parameters - mean=-3.25, standard deviation=.9


                               | Samp_size          Perf         SE   CI width
-------------------------------+----------------------------------------------
              Criteria 1 - O/E |      7305             1       .051         .2
          Criteria 2 - C-slope |     11307             1       .051         .2
      Criteria 3 - C statistic |      1967           .74       .026         .1
-------------------------------+----------------------------------------------
                      Final SS |     11307             1       .051         .2

Minimum sample size required for model validation based on user inputs = 11307,
with 566 events (assuming an outcome prevalence = .05)

Criteria 1 - precise estimation of O/E performance in the validation sample
Criteria 2 - precise estimation of the calibration slope in the validation sample
Criteria 3 - precise estimation of the C statistic in the validation sample
```

```
. pmvalsampsize, type(b) prevalence(0.05) cstat(0.74) lpnormal(-3.25, 0.9) graph
Normal LP distribution with parameters - mean=-3.25, standard deviation=.9
```

	Samp_size	Perf	SE	CI width
Criteria 1 - O/E	7305	1	.051	.2
Criteria 2 - C-slope	11307	1	.051	.2
Criteria 3 - C statistic	1967	.74	.026	.1
Final SS	11307	1	.051	.2

```
Minimum sample size required for model validation based on user inputs = 11307,
with 566 events (assuming an outcome prevalence = .05)

Criteria 1 - precise estimation of O/E performance in the validation sample
Criteria 2 - precise estimation of the calibration slope in the validation sample
Criteria 3 - precise estimation of the C statistic in the validation sample
```

```
. pmvalsampsize, type(b) prevalence(0.05) cstat(0.74) lpnormal(-3.25, 0.9) graph
Normal LP distribution with parameters - mean=-3.25, standard deviation=.9
```

	Samp_size	Perf	SE	CI width
Criteria 1 - O/E	7305	1	.051	.2
Criteria 2 - C-slope	11307	1	.051	.2
Criteria 3 - C statistic	1967	.74	.026	.1
Final SS	11307	1	.051	.2

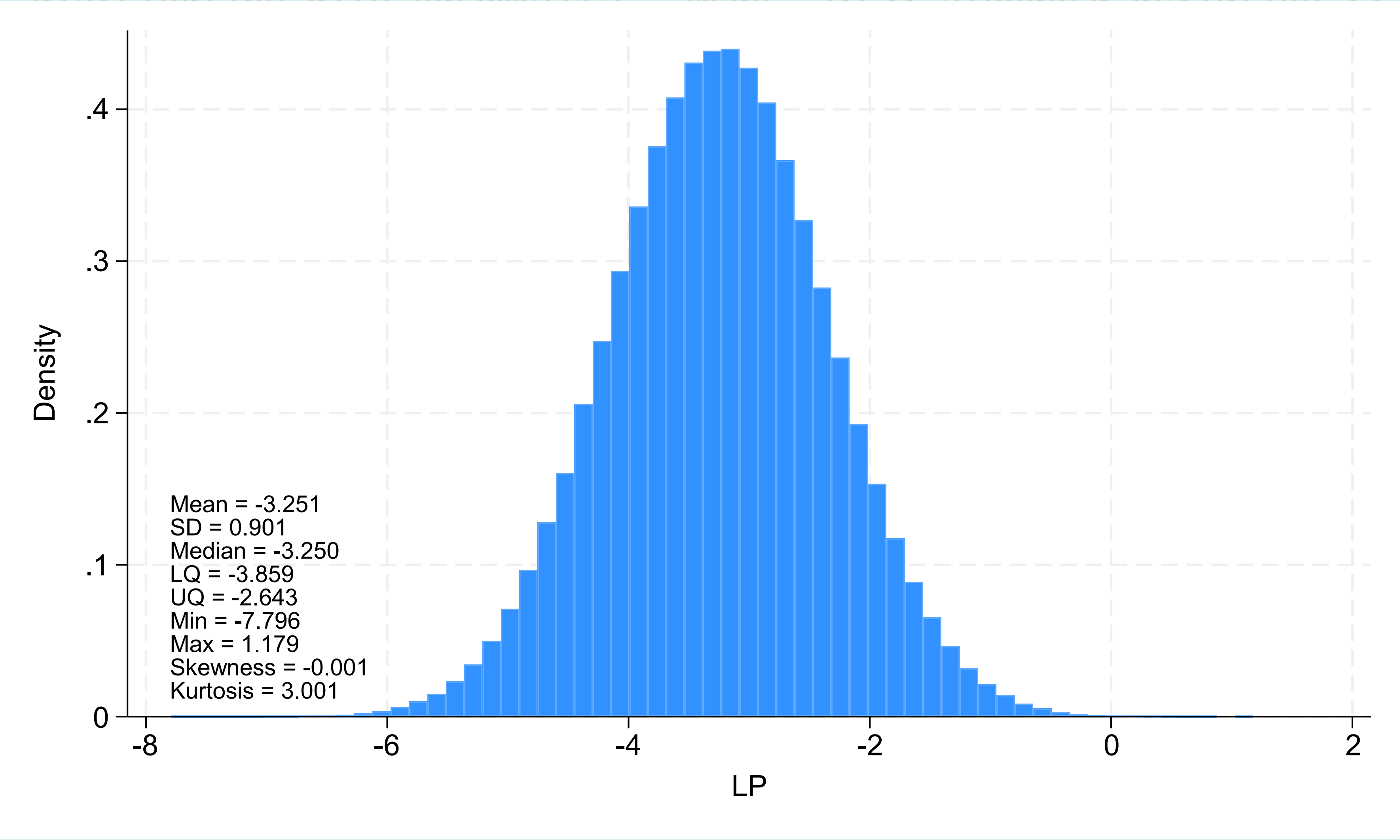
```
Minimum sample size required for model validation based on user inputs = 11307,
with 566 events (assuming an outcome prevalence = .05)

Criteria 1 - precise estimation of O/E performance in the validation sample
Criteria 2 - precise estimation of the calibration slope in the validation sample
Criteria 3 - precise estimation of the C statistic in the validation sample
```

```
. pmvalsampsize, type(b) prevalence(0.05) cstat(0.74) lpnormal(-3.25, 0.9) graph
```

- Options to specify LP distribution include:
  - Normal
  - Skewed normal
  - Beta – for predicted probabilities
  - C statistic based normal distributions

- Graph option for checking

```
. pmvalsampsize, type(b) prevalence(0.05) cstat(0.74) lpnormal(-3.25, 0.9) graph
Normal LP distribution with parameters - mean=-3.25, standard deviation=.9
```



```
. pmvalsampsize, type(b) prevalence(0.05) cstat(0.74) lpcstat(-3.3) graph
Proportion of observed outcome events is within tolerance
Proportion of outcome events under simulation = .053031 + Target prevalence = .05
Mean in non-event group=-3.3
```

	Samp_size	Perf	SE	CI width
Criteria 1 - O/E	[illegible]	1	.051	.2
Criteria 2 - C-slope	10454	1	.051	.2
Criteria 3 - C statistic	1967	.74	.026	.1
Final SS	10454	1	.051	.2

```
Minimum sample size required for model validation based on user inputs = 10454,
with [illegible] events (assuming an outcome prevalence = .05)

Criteria 1 - precise estimation of O/E performance in the validation sample
Criteria 2 - precise estimation of the calibration slope in the validation sample
Criteria 3 - precise estimation of the C statistic in the validation sample
```

- Specify C statistic & non-event mean starting value
- Options to aid iteration process include:
  - Trace
  - Tolerance
  - Iteration step
- Strong assumptions

"Reported model performance measures:

Discrimination = 57/78 (73%)

Calibration = 11/78 (14%)

Overall metrics = 18/78 (23%)"

Collins et al. 2014

pmstats

- Many proposed performance statistics exist
  - Time consuming & confusing

- R users have `rms`

- pmstats calculates key performance measures including:
  - Discrimination
  - Calibration
  - Overall performance
  - Reporting statistics

Predictions in a new sample

- Assuming we have the full published heart surgery model of the form:

$$logit(p) = LP = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ \ldots$$

- Manually generate a new LP variable

```
gen LP =  -4.55 + (.49*sex) + (.0073*age) + (2.48*histDiabetes) +
(1.46*histMI) + (.67*histCVA) + (.37*histPCI) + …
```

- Given `LP` & `outcome` we can now assess the models external performance

pmstats

- Estimates with CI’s
- Calibration model parameters
- Continuous & TTE outcomes

Discrimination statistics ...

	Estimate	SE	Lower_CI	Upper_CI
C-Statistic	0.765	0.043	0.681	0.848
Somers D	0.529	0.085	0.362	0.697

Calibration statistics ...

	Estimate	Lower_CI	Upper_CI
O/E	0.228	0.001	48.229
E-O	0.399	0.347	0.438
CITL	-2.700	-3.081	-2.320
C-Slope	0.840	0.487	1.193

Further information

Overall performance statistics ...

	Estimate	Lower_CI	Upper_CI
Cox-Snell R2	0.096	0.051	0.170
R2 Nagelke~e	0.185	0.102	0.310
R2 McFadde~s	0.138	0.075	0.240
Briers Score	0.287	0.259	0.319

Additional summary statistics ...

	Mean	SD	Median	LQ	UQ	Min	Max	Skewness	Kurtosis
LP Dist	0.047	1.449	0.181	-0.979	1.189	-3.946	3.295	-0.342	2.468
Sample size	296.000	.	.	.	.	.	.	.	.
Events	35.000	.	.	.	.	.	.	.	.

- Overall performance statistics
- Linear predictor distribution useful for future research

“Only 11 studies presented a calibration plot (11/78; 14% 95% CI 8% to 24%)”

Collins et al. 2014

pmcalplot

- Using the same validation sample

- Predicted probabilities calculated using LP

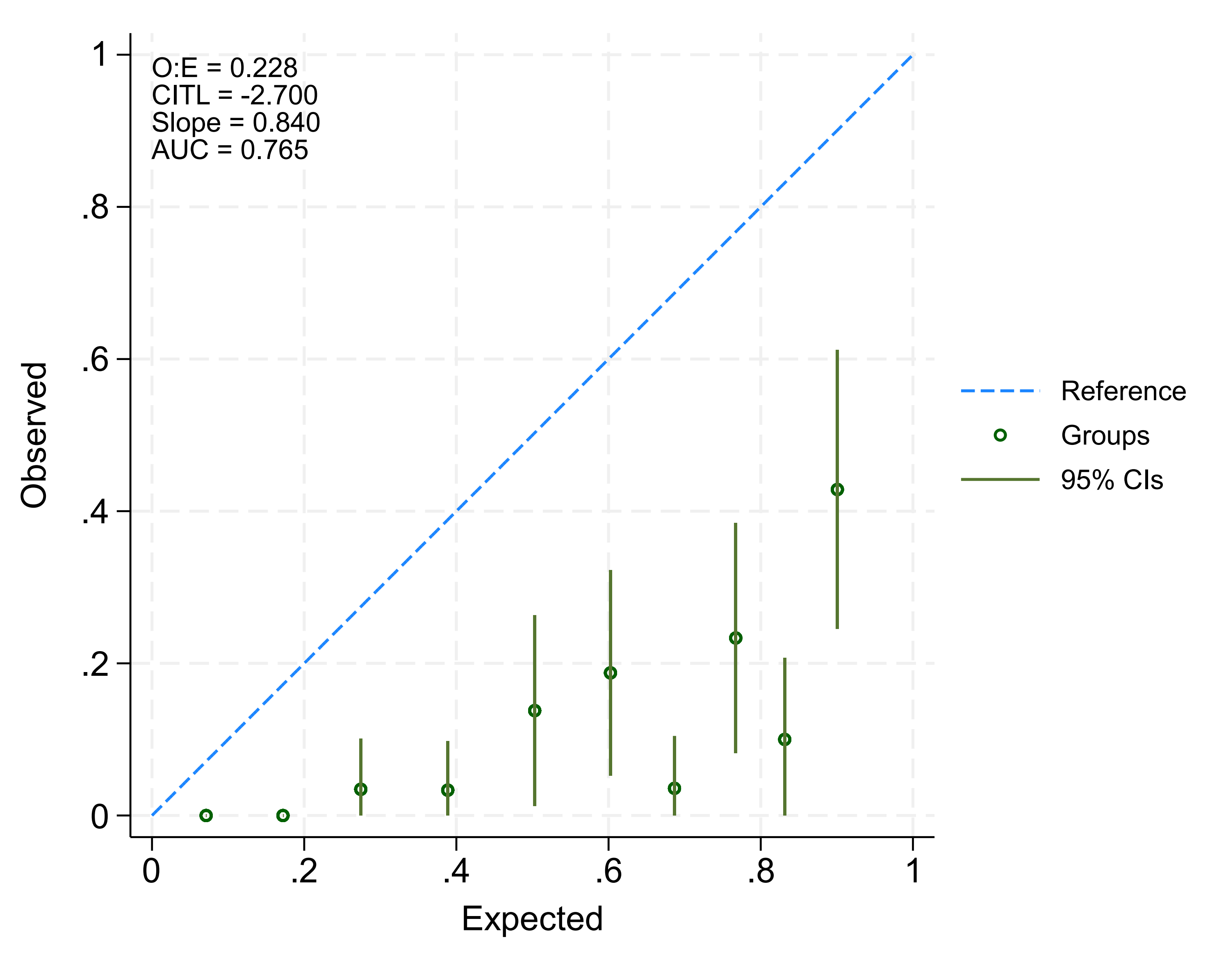
$$p = \frac{e^{LP}}{1 + e^{LP}}$$

- Generate the predicted probabilities

```
gen pr = (exp(LP))/(1+exp(LP))

pmcalplot pr outcome, ci
```

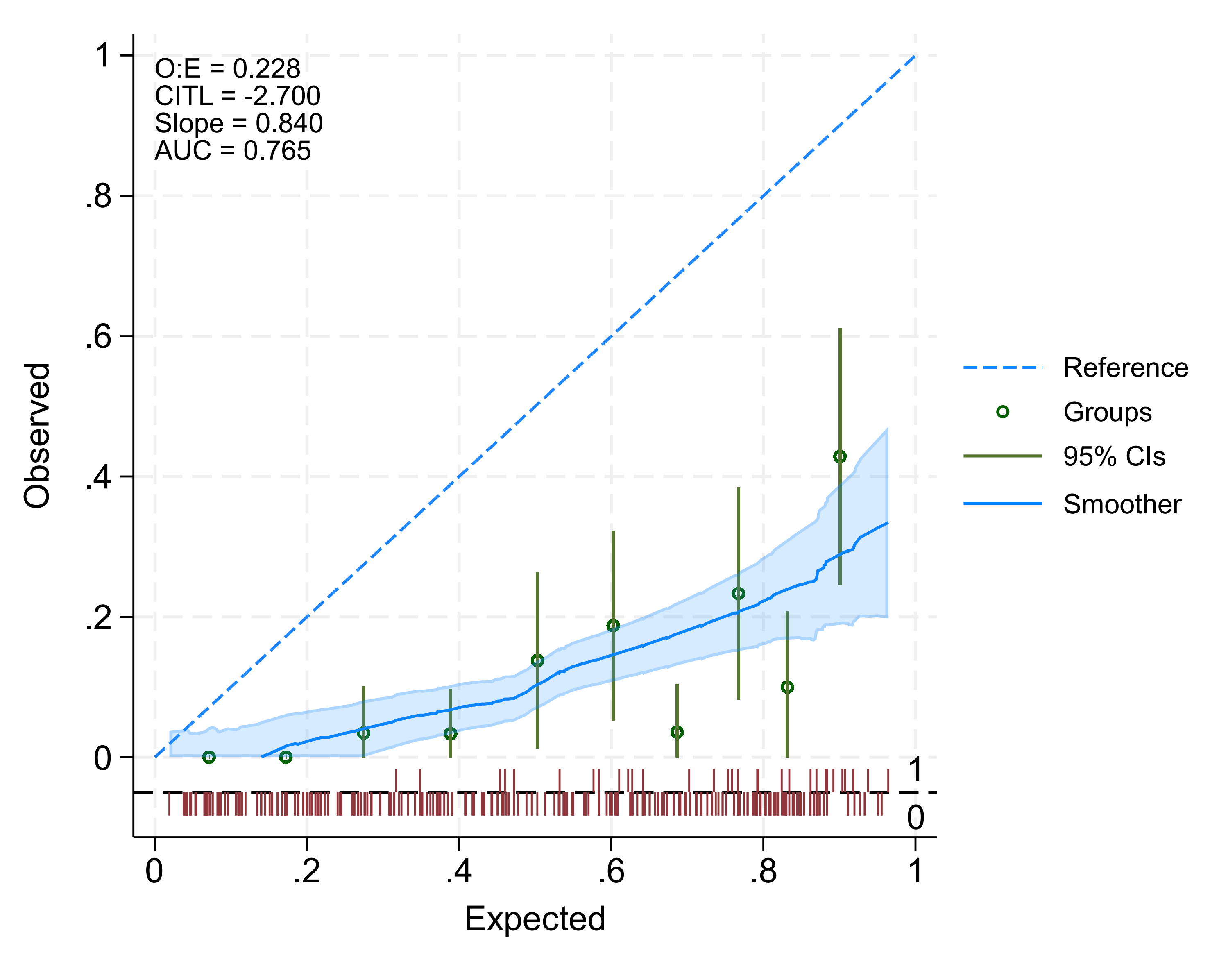
External validation

- Plots observed outcomes against predictions from the model
- Historically plotted in groupings

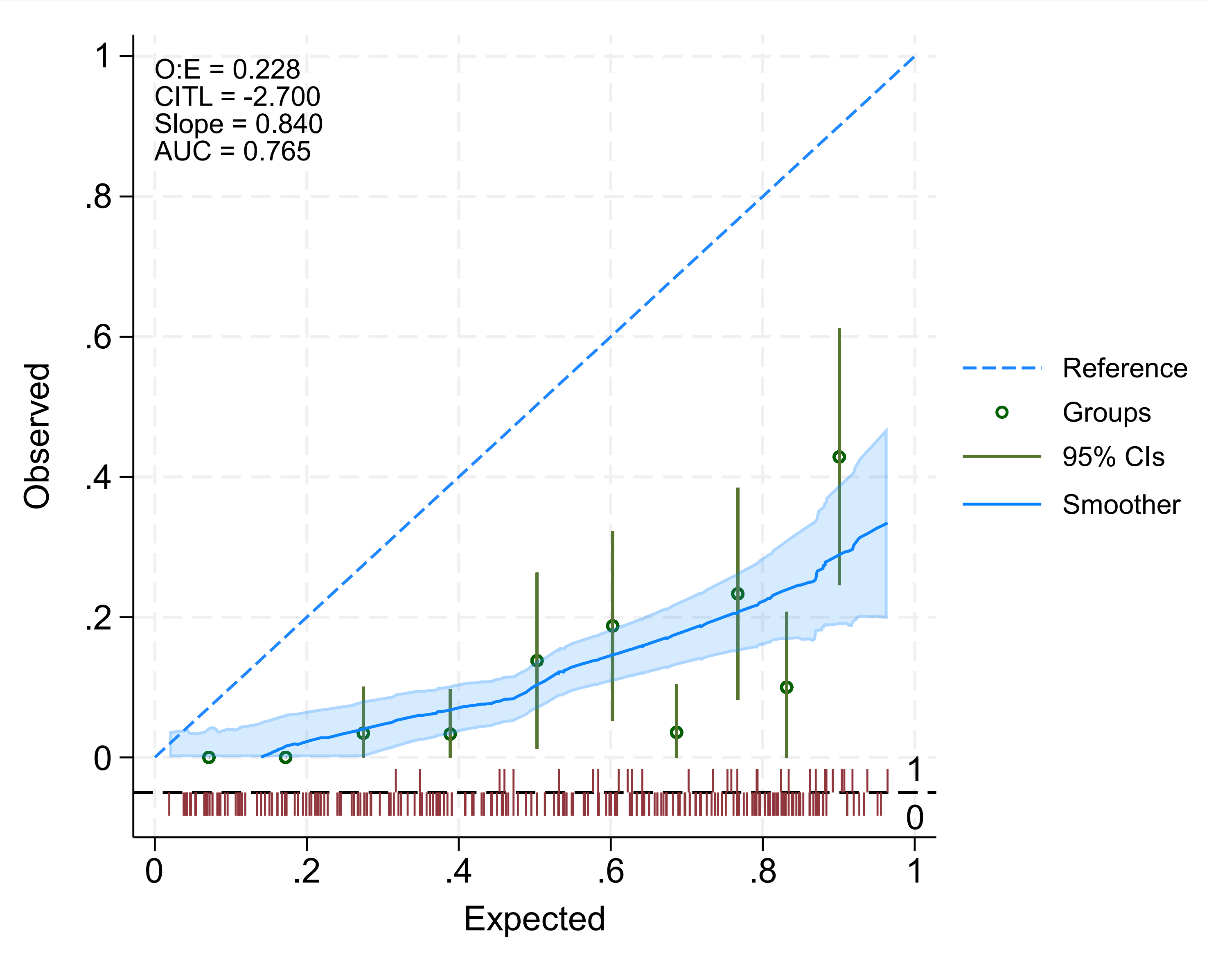


- Calibration curve allows assessment of calibration at the individual patient level

- Spike plot showing the spread of events/non-events across risk spectrum

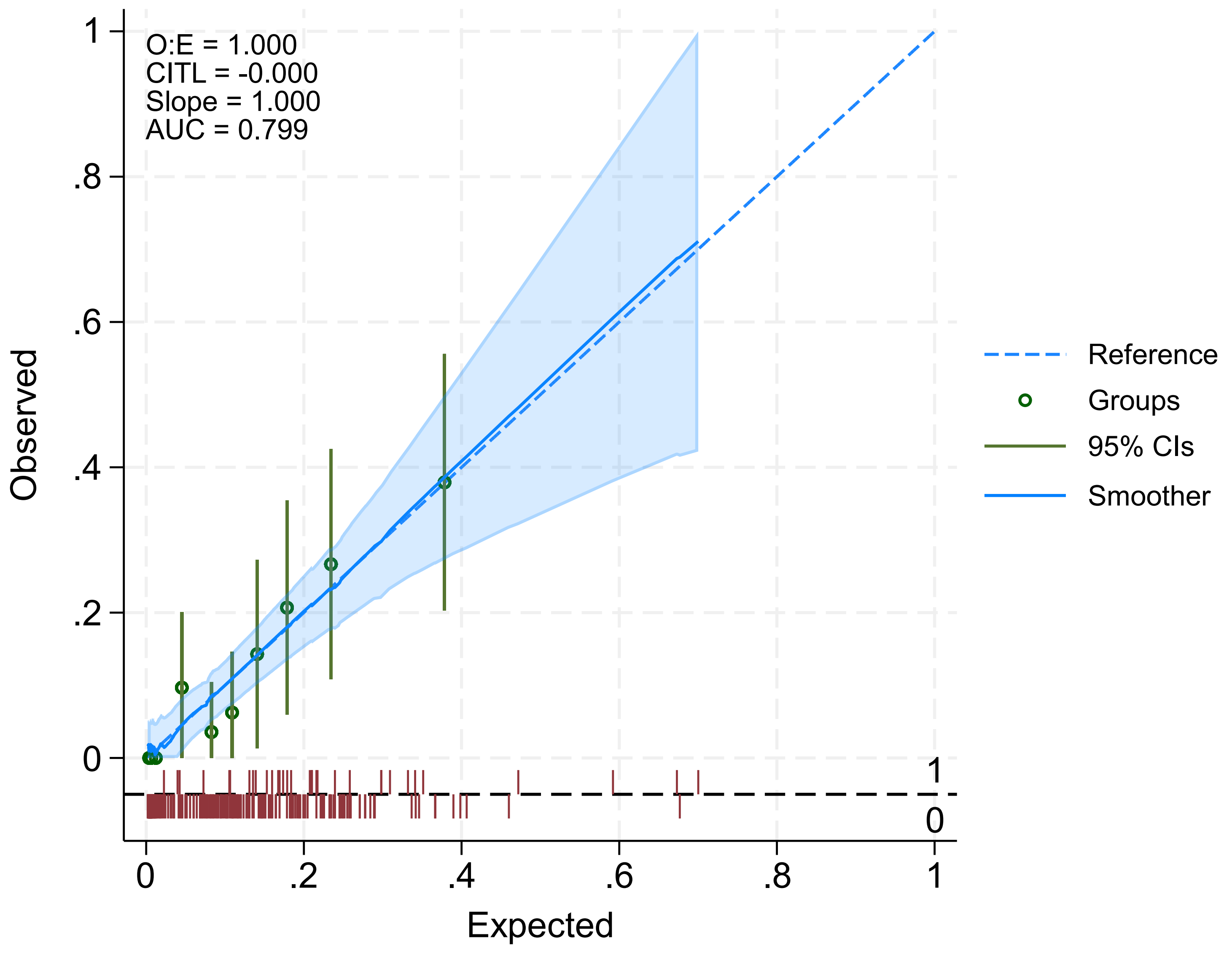


- Clear over-prediction
- Systematic miscalibration
- Evidence of overfitting



Apparent performance

- Primarily for external validation
- Can be used to check apparent performance!



Final thoughts

pm-suite

Development & Internal Validation → **External Validation & Updating**

pmsampsize
pmintval

pmvalsampsize
pmupdate

pmstats
pmcalplot

pmiecv
pmcstat

pmsplot
pmmeta

Take home

- Important to describe your sample size
  - number of events
  - number of candidate predictor parameters
  - how you came up with your sample size

- Multiple measures of model performance
- Calibration plots
- Baseline survival/hazard at multiple time-points
- Distribution of linear predictor
- Range of predictors

Conclusion

- Prediction modelling is hard!
  - Easy to end up with inaccurate and unreliable models

- Carefully consider
  - Design
  - Evaluation

- Fully report all stages

With thanks to Richard Riley, Gary Collins, Kym Snell, Lucy Archer …

Thank *you*

Email: j.ensor@bham.ac.uk
Twitter: @joie_ensor

UNIVERSITY OF BIRMINGHAM