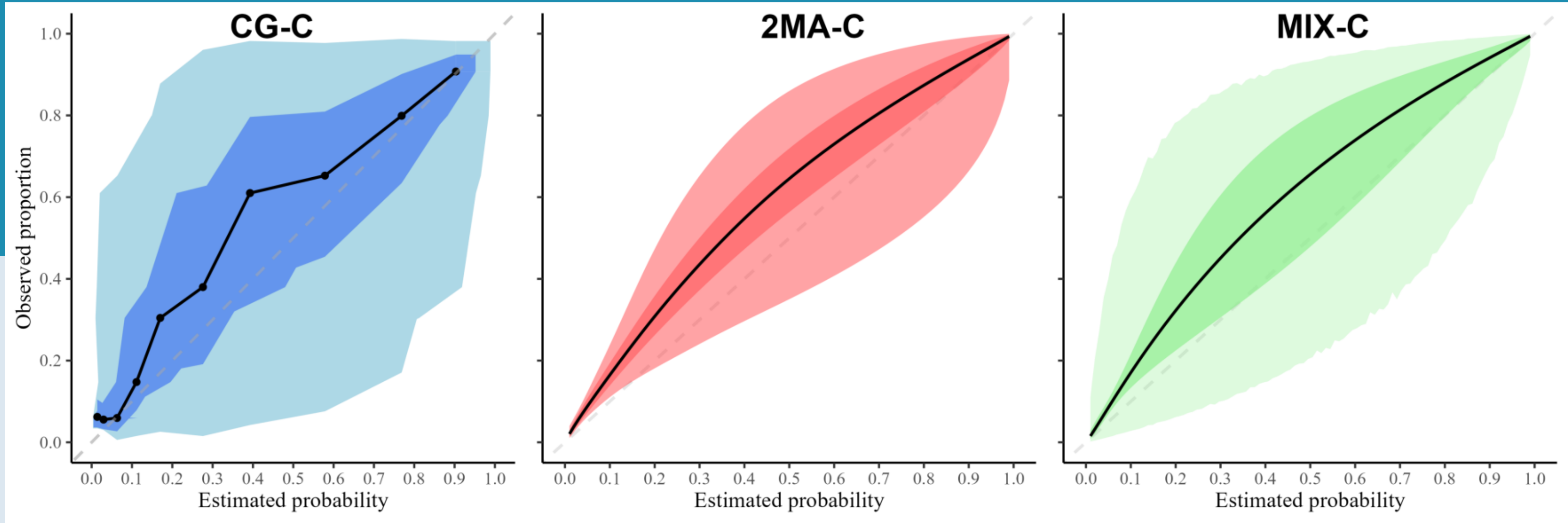
30/04/2025, MEMTAB, Birmingham

KU LEUVEN

Combining **calibration** plots from multiple **clusters**

A phase 2 methodological study

Lasai Barreñada, Bavo D.C. Campo,
Laure Wynants & Ben Van Calster

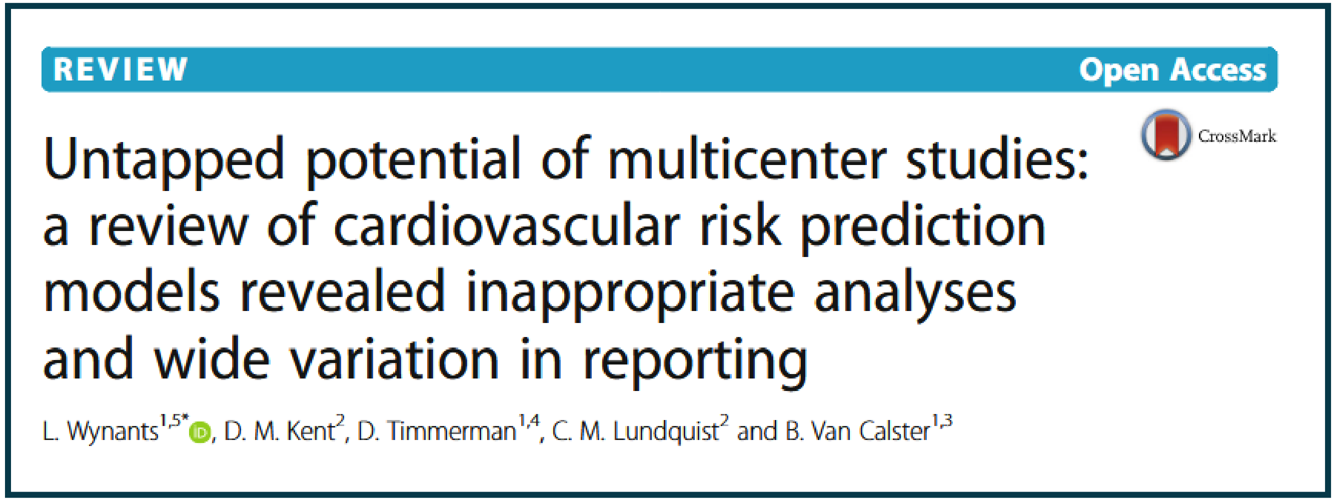


fwo

LUHTAR
LEUVEN UNIT FOR HTA RESEARCH

Clustered clinical data

- **Examples**:
  - Studies in a meta-analysis
  - Centers in a multicenter study

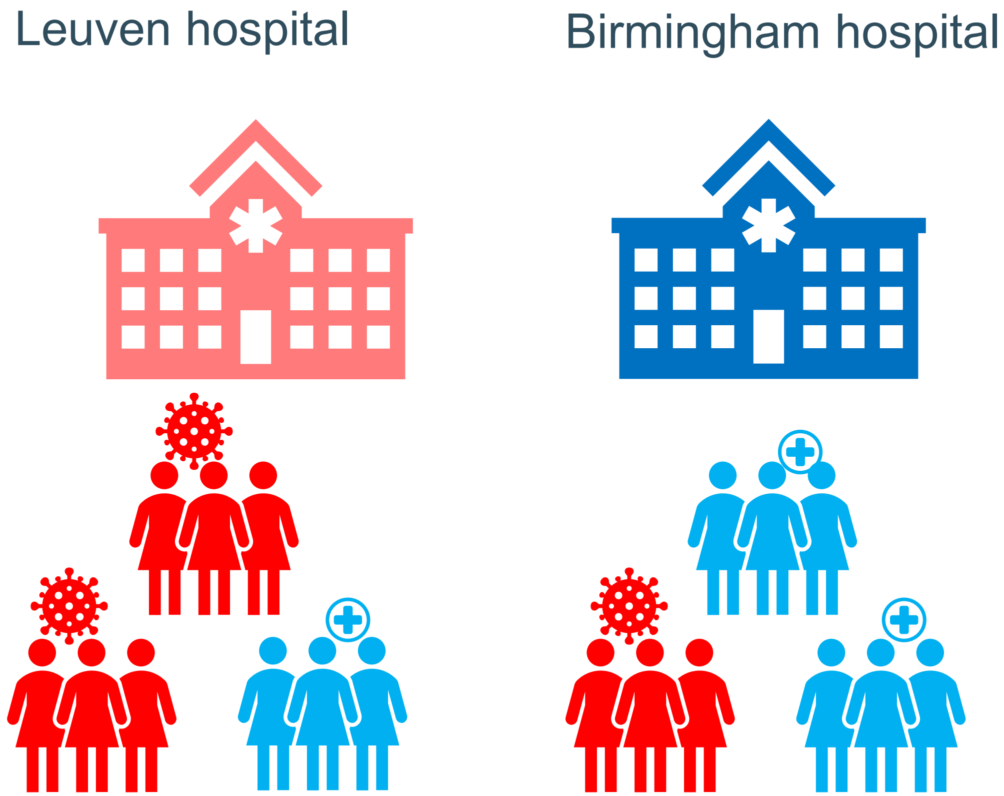


REVIEW — Open Access

CrossMark

Untapped potential of multicenter studies: a review of cardiovascular risk prediction models revealed inappropriate analyses and wide variation in reporting

L. Wynants[1,5*], D. M. Kent[2], D. Timmerman[1,4], C. M. Lundquist[2] and B. Van Calster[1,3]

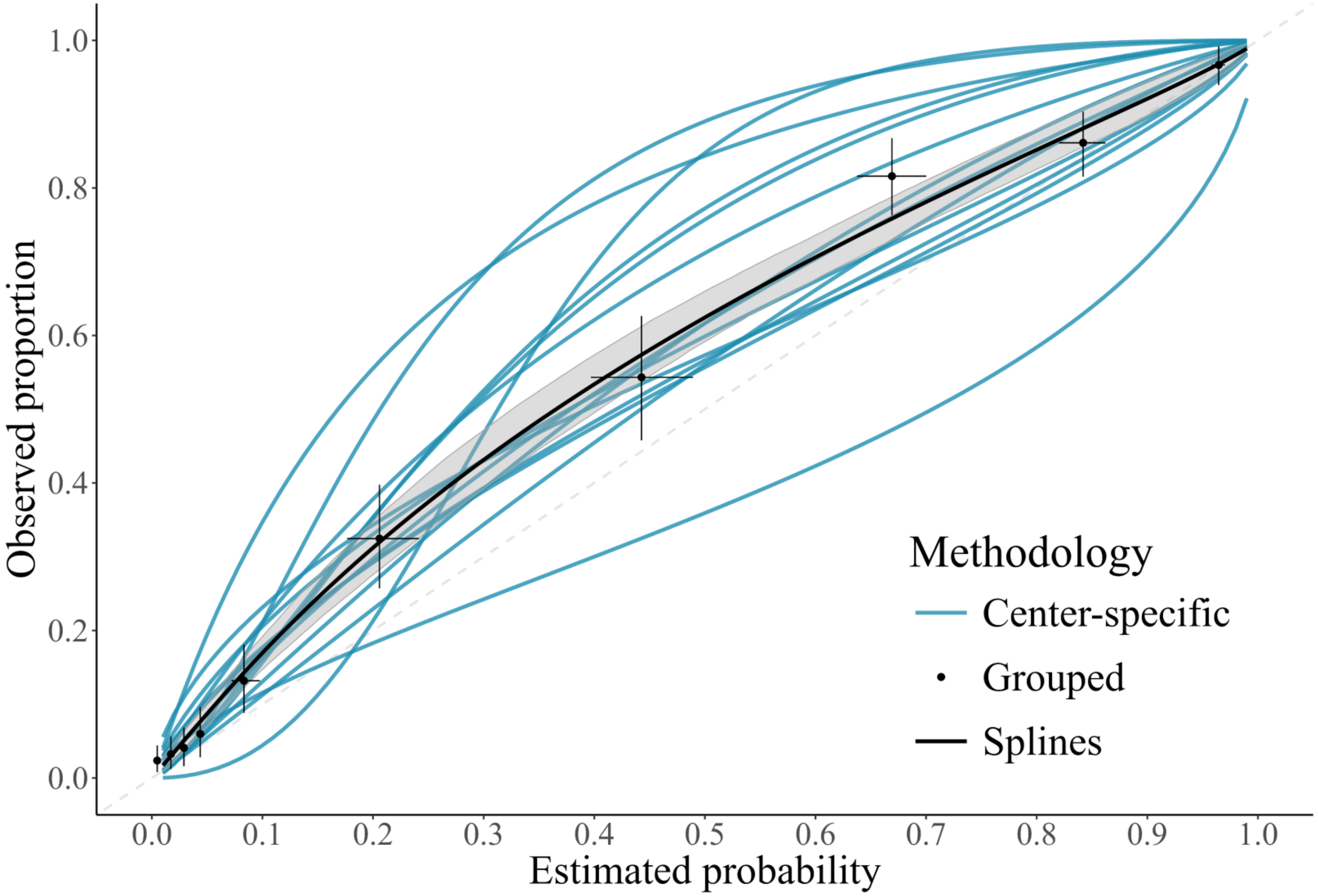


Wynants et al. Diagnostic and Prognostic Research (2019) 3:6

Evaluation of calibration

- **Logistic calibration framework**
  - $logit(Y_{ij}) = \alpha + \beta LP$.
  - Flexible: LOESS, Splines...

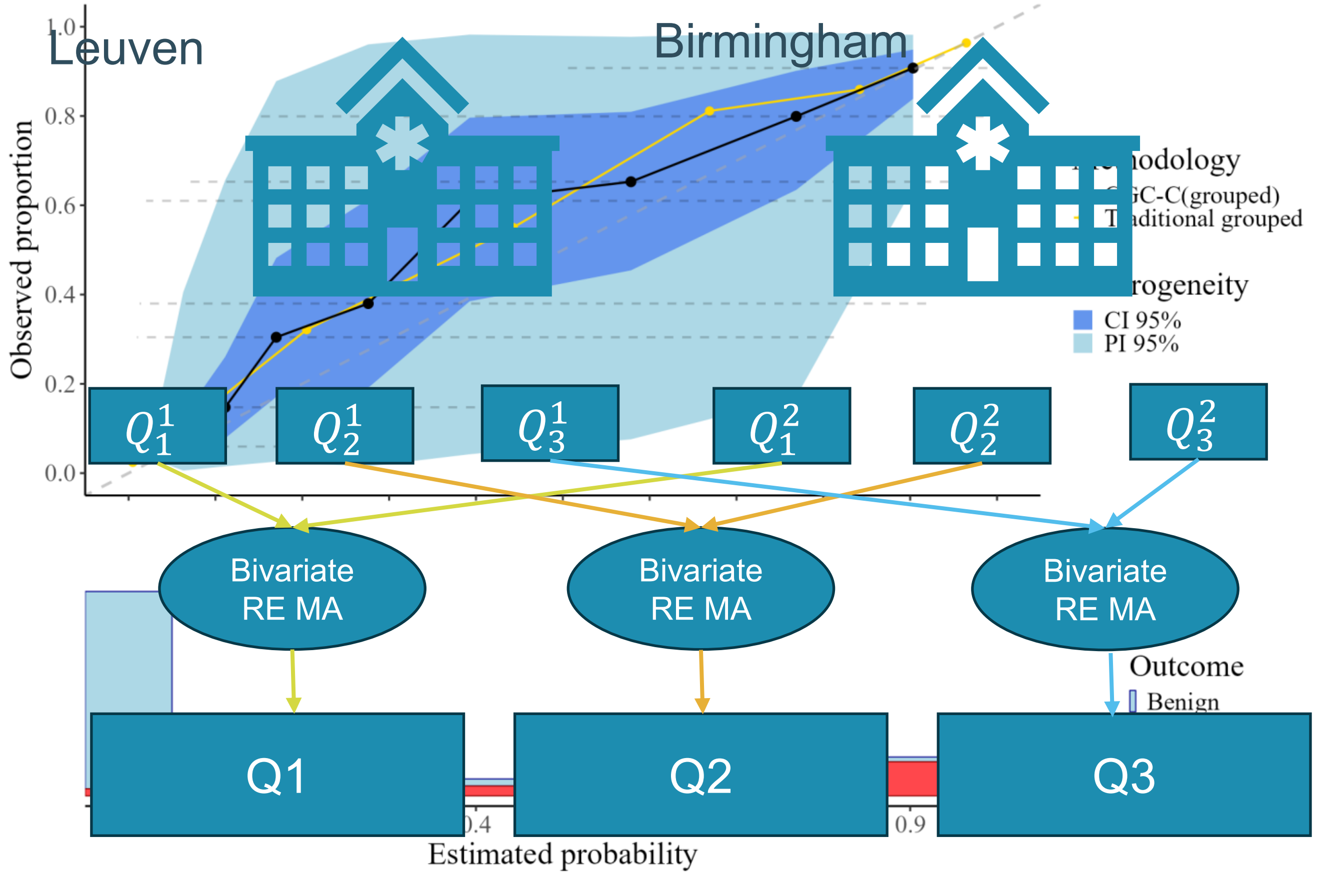
- **Calibration plot** is most informative assessment
  - X-axis: Predicted risks
  - Y-axis: Observed proportion
  - Can include 95% CI
- **Illustration**: Ovarian cancer data
  - N = 2489
  - 14 clusters (size 38 to 369)



KU LEUVEN

Clustered group calibration (CG-C)

Ovarian cancer data



$$PI = \hat{y}_q \mp t_{J-2}\sqrt{\tau_y^2 + SE(\hat{y}_q)^2}$$

1. Group the predicted probabilities by center in Q groups → Quantile or interval
2. Calculate prevalence and mean predicted risk per group.
3. Meta-analyse each centre's groups
4. Calculate confidence and prediction intervals

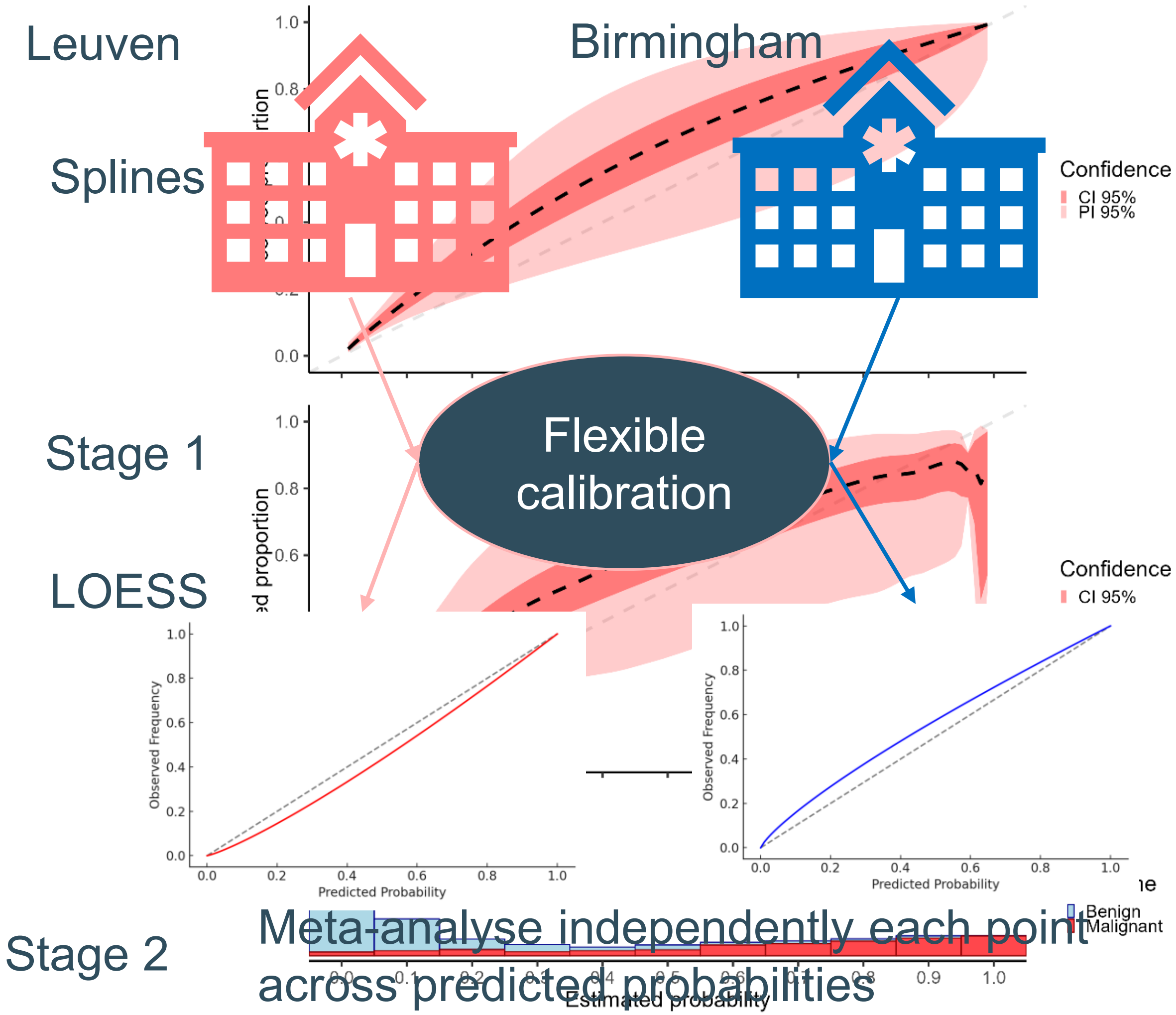
- **Model agnostic center specific approach.**
- **Easy to compute and explain.**
- **Very dependant on number of quantiles.**
- **Not possible to obtain center-specific curves.**

Riley et al., BMJ, 2011

KU LEUVEN

Two stage meta-analysis calibration (2MA-C)

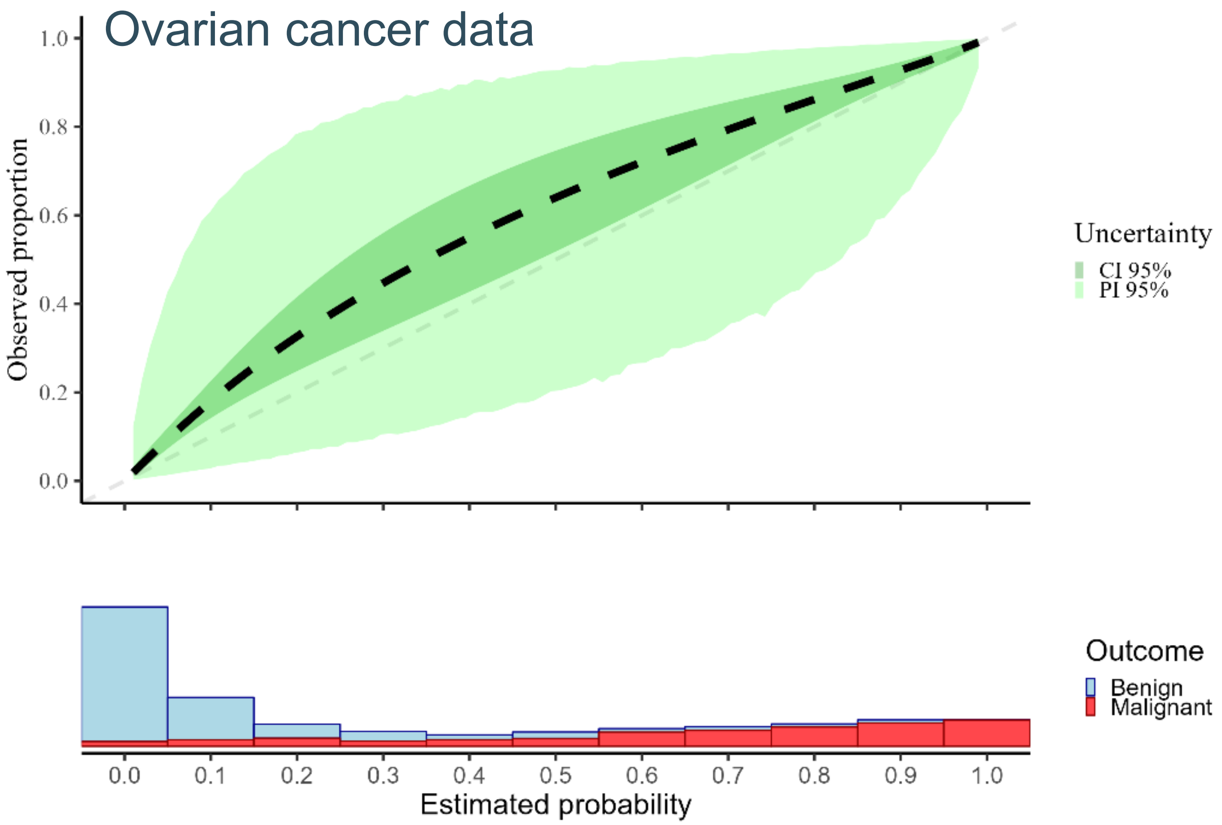
Ovarian cancer data



1. Fit center specific flexible calibration models.
2. Use fitted model to predict center specific observed proportions for a grid of values.
3. Meta-analyse each point in the grid per center.
4. Calculate confidence and prediction intervals.

- **Confidence and prediction intervals based on RE meta-analysis.**
- **Depends on the model fitted in each center.**
- **Does not estimate center specific curves.**
- **Models for each point in the grid are independent**
- **Computationally costly (~200x)**

Mixed model calibration (MIX-C)

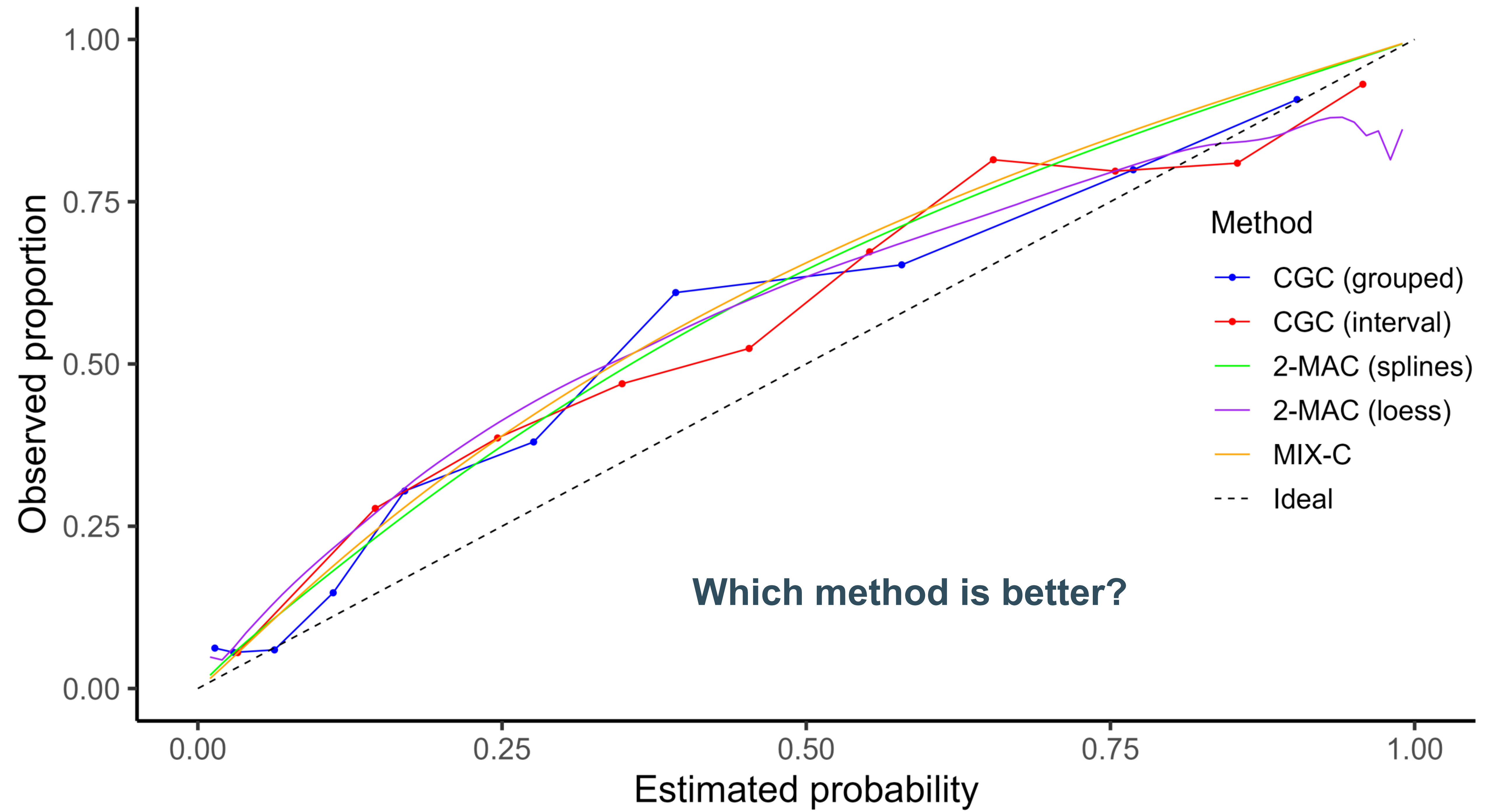


$$\hat{Y}_{ij} = a + b\boldsymbol{f}(x)_{ij} + R_{ij}$$

- Logistic mixed effects model (GLME) with random slopes and intercepts per center.
- Simulation based confidence intervals.

- **Center specific curves accounting for clustering**
- **Not model agnostic.**
- **Computationally costly (~400x).**

KU LEUVEN



Observed proportion
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Estimated probability
Method
CGC (grouped)
CGC (interval)
2-MAC (splines)
2-MAC (loess)
MIX-C
Ideal
Which method is better?

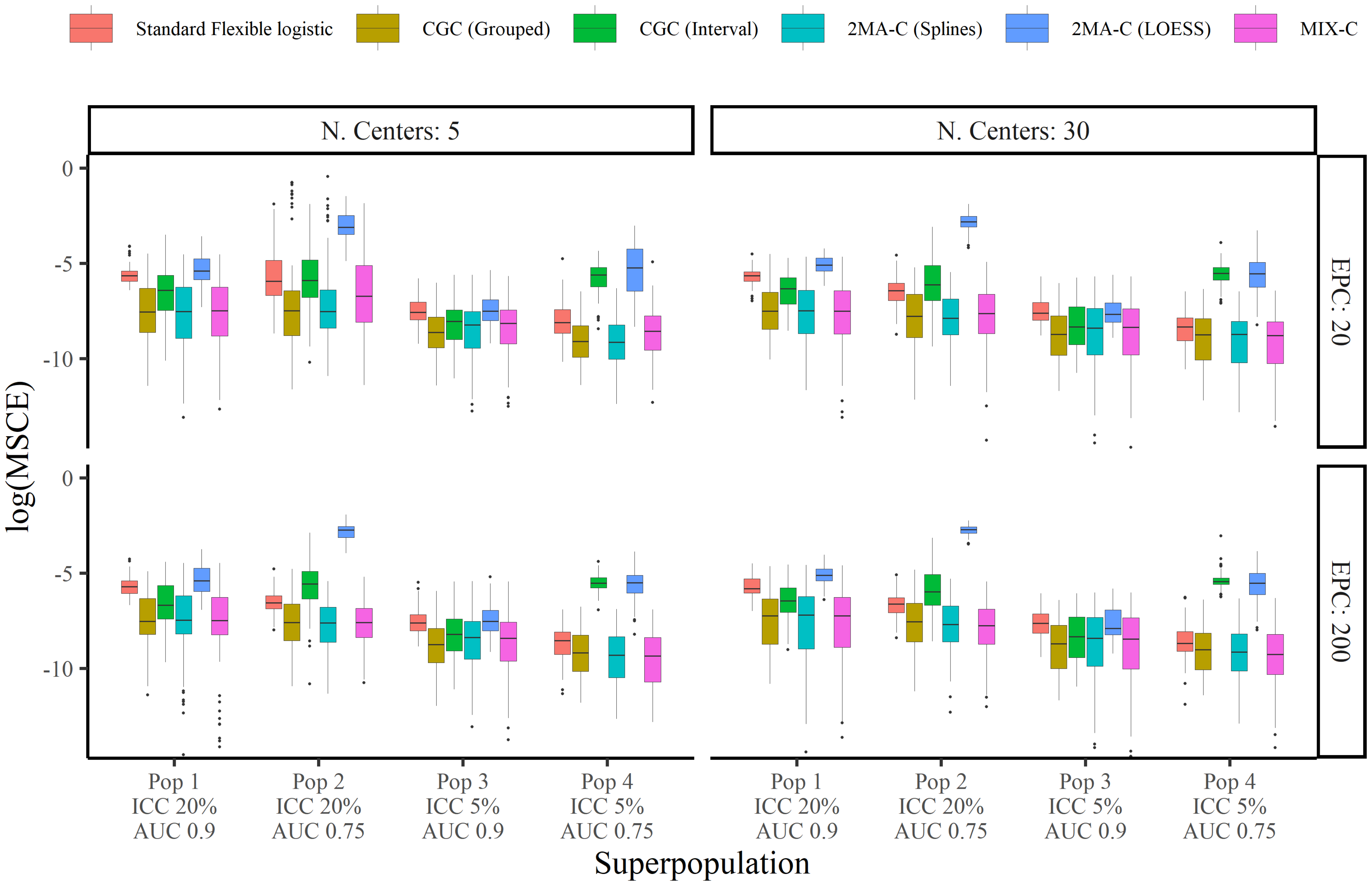
Simulation study

Superpopulation	AUC	ICC	ER	Formula
P1	0.9	0.2	0.3	logit(p) = -1.605469 + -2.0906250X + res(0,1.5593722)
P2	0.75	0.2	0.3	logit(p) = -1.012207 + 0.4199219X + res(0,1.0024963)
P3	0.9	0.05	0.3	logit(p) = -1.594375 + 2.3875000X + res(0, 0.7827540)
P4	0.75	0.05	0.3	logit(p) = -1.024414 + --0.9273437X + res(0, 0.5183335)

- Each superpopulation has 200 centers and 10000 observations per center.
- From them take samples with different EPC (20, 200) and number of centers (5, 30)
- Train a logistic regression model and evaluate in a big sample.
- **Mean square calibration error (MSCE):** mean square difference of the true average risk (aj = 0) and the estimated observed proportion
- **Prediction interval coverage**: center specific observed proportion included in the PI.

Results (MSCE)

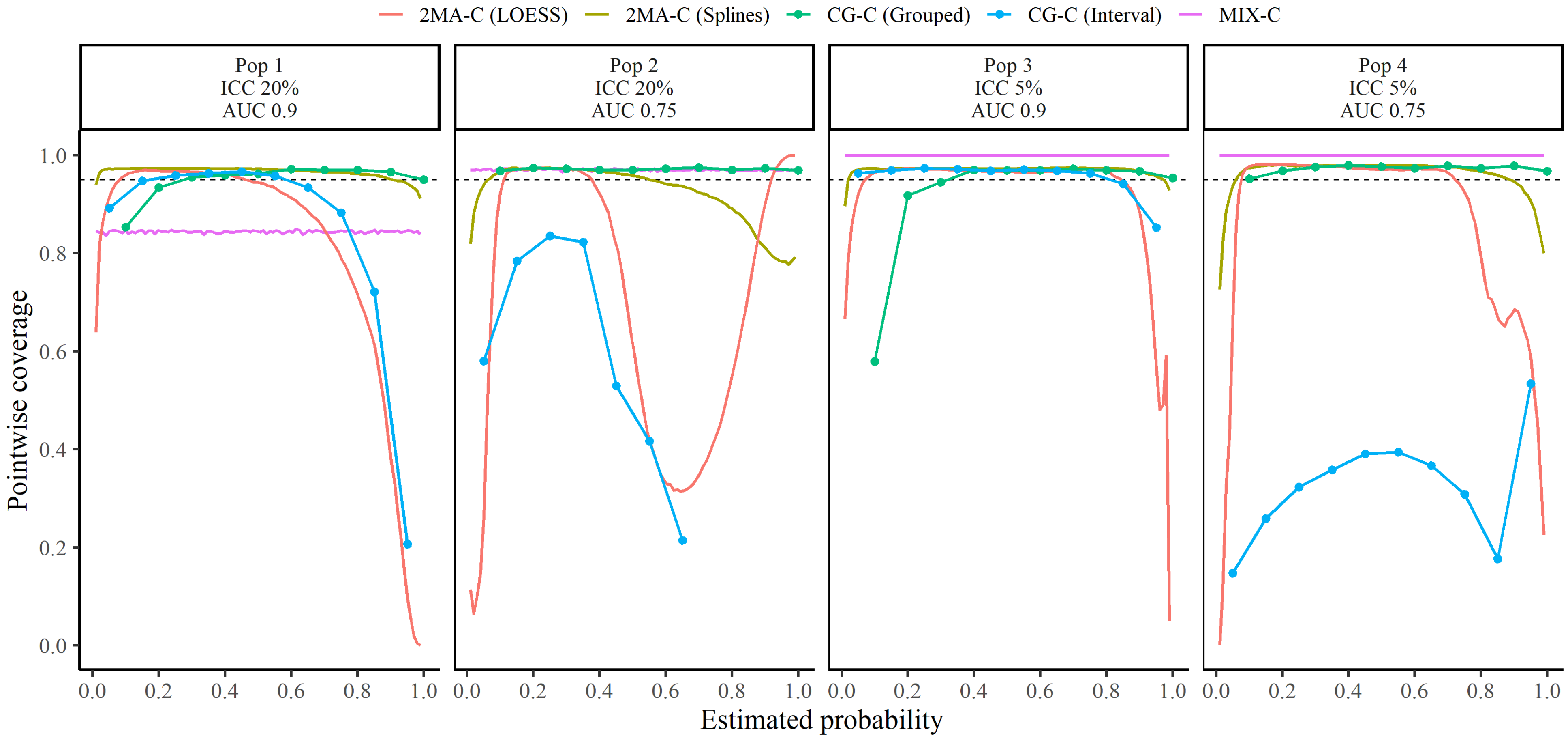
- **2MA-C (splines) and MIX-C**
- **CGC (interval) and LOESS**
- **Truth is LR**



KU LEUVEN

Results (PI coverage)

- 2MA-C (splines) the best method.
- None of the methods correctly estimate the PI.



Take home messages

- Taking clustering into account when evaluating calibration is important
- All methods have suboptimal prediction interval coverage
- We recommend 2MA-C (splines) for the average curve and MIX-C for center specific clustered calibration (results not shown in this presentation)
- Ready to use code and functions available and soon to be incorporated to CalibrationCurves R package

Preprint link

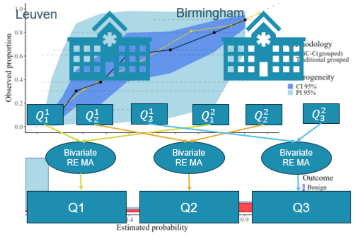
Thank you!

Method	Estimation of observed proportion	Strengths	Limitations
CG-C	**Grouped**: Bivariate random effects meta-analysis of logit-transformed mean estimated risk and event fraction by quantile per cluster. **Interval**: Bivariate random effects meta-analysis of logit-transformed mean estimated risk and event fraction by estimated risk interval.	+ Model agnostic + Pointwise confidence and prediction intervals. + All clusters have the same number of groups	- Computation time - Groups can contain observations with very different estimated risks within and between clusters (Grouped version). - Clusters may not have the same number of groups (e.g. risk intervals without observations). (Interval version) - Curves depend on number of groups
2MA-C	Random effects meta-analysis of estimated smooth observed proportion by cluster **Splines**: Recommended when clusters are small **LOESS**: More flexible but can fail with small clusters.	+ Pointwise confidence and prediction intervals.	- Computation time - Curve dependent on the smoother used in the cluster-specific models.
MIX-C	Logistic generalized linear mixed model with restricted cubic splines.	+ Curvewise confidence and prediction intervals. + Provides also shrunken curves per center.	- Computation time

KU LEUVEN

Clustered group calibration (CG-C)

Ovarian cancer data



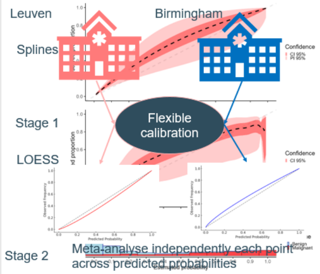
$$PI = \hat{y}_q \mp t_{J-2}\sqrt{\tau_y^2 + SE(\hat{y}_q)^2}$$

1. Group the predicted probabilities by center in Q groups → Quantile or interval
2. Calculate prevalence and mean predicted risk per group.
3. Meta-analyse each centre's groups
4. Calculate confidence and prediction intervals

- **Model agnostic center specific approach.**
- **Easy to compute and explain.**
- **Very dependant on number of quantiles.**
- **Not possible to obtain center-specific curves.**

Two stage meta-analysis calibration (2MA-C)

Ovarian cancer data

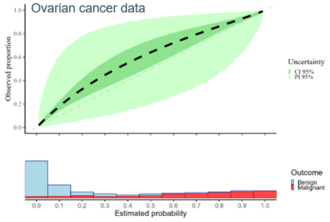


1. Fit center specific flexible calibration models.
2. Use fitted model to predict center specific observed proportions for a grid of values.
3. Meta-analyse each point in the grid per center.
4. Calculate confidence and prediction intervals.

- **Confidence and prediction intervals based on RE meta-analysis.**
- **Depends on the model fitted in each center.**
- **Does not estimate center specific curves.**
- **Models for each point in the grid are independent**
- **Computationally costly (~200x)**

Mixed model calibration (MIX-C)

Ovarian cancer data



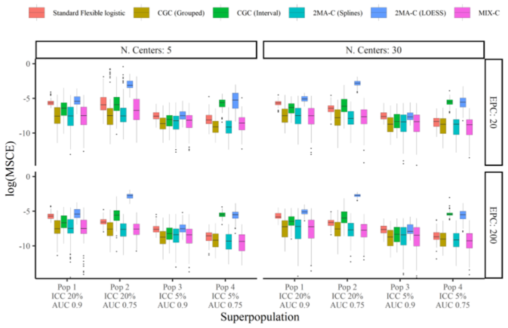
$$\hat{Y}_{ij} = a + b\boldsymbol{f}(x)_{ij} + R_{ij}$$

- Logistic mixed effects model (GLME) with random slopes and intercepts per center.
- Simulation based confidence intervals.

- **Center specific curves accounting for clustering**
- **Not model agnostic.**
- **Computationally costly (~400x).**

Results (MSCE)

- **2MA-C (splines) and MIX-C**
- **CGC (interval) and LOESS**
- **Truth is LR**



Results (PI coverage)

- 2MA-C (splines) the best method.
- None of the methods correctly estimate the PI.