SVEPM 2017

Diagnostic models applied in new clusters With informative priors for the random effects

Utrecht University

Haifang Ni, Mirjam Nielen, Rolf H.H. Groenwold, Irene Klugkist

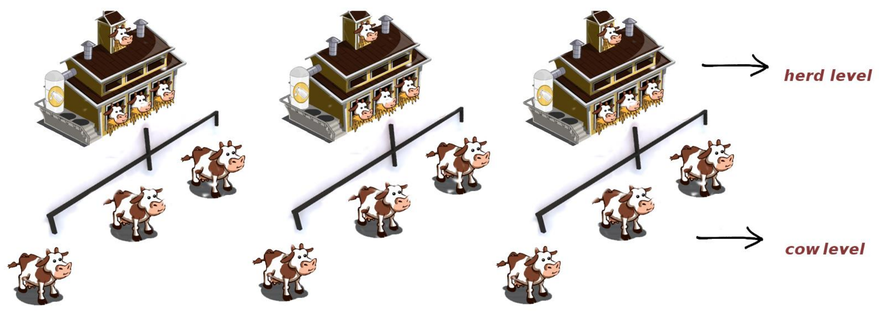
Contact info: h.ni@uu.nl

Background and objective

Clustering is often seen in clinical practice. **Random effects regression models** are routinely used for clustered data.

For **diagnostic (prediction)** models with clustered data, at the model development phase, random effects regression models are also seen to be used. However, after the model is developed and parameter estimates determined, **the random effect term is typically left out** and predictions **for future new clusters** are done ignoring the between cluster variance (e.g., see reference 1,2).

This *simulation study* Shows that we **can keep** the random effects in the diagnostic model when applied in new clusters. Results also show that by adding **cluster level** simulated **expert knowledge as priors for the random effects**, the diagnostic accuracy improves.



Methods

A simple multilevel logistic regression model is used where there is one subject level predictor and one intercept. The intercept contains a fixed part β_0 and a random part u_j (i.e., the random cluster effect).

$$logit\big(p(y_{ij}=1)\big) = \beta_0 + \beta_1 x_{1ij} + u_j$$
$$\pi(u_j) \propto N(0, \sigma_u^2)$$

Using model development data, we get parameter estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_u^2$.

1. How do we keep the random effects in the diagnostic model?

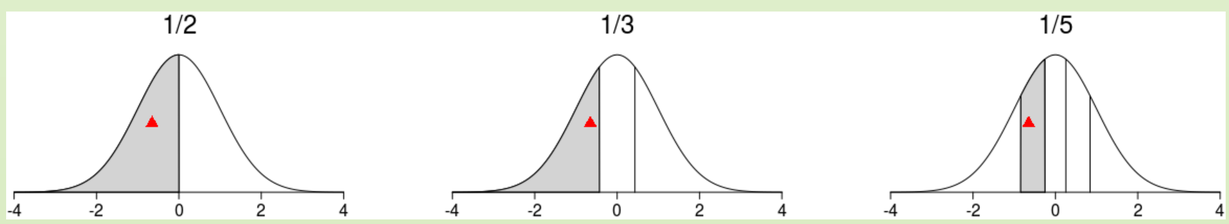
By using **Bayesian** modeling, we sample a value from the distribution $N(0, \hat{\sigma}_u^2)$ as the predicted random effect $\hat{u}_c$ **for new cluster** c.

2. Do predictions improve with random effects attained in the model?

No, it **adds uncertainty** to the prediction which is more realistic. But diagnostic accuracy is similar to the model with random effects removed.

3. How can we improve predictions using the random effects model?

By **adding cluster specific prior** information **to predict the random effect** $\hat{u}_c$. Hence, instead of sampling from the whole distribution $N(0, \hat{\sigma}_u^2)$, we sample a value for $\hat{u}_c$ from **a truncated area of the distribution** $N(0, \hat{\sigma}_u^2)$ (see Figure).



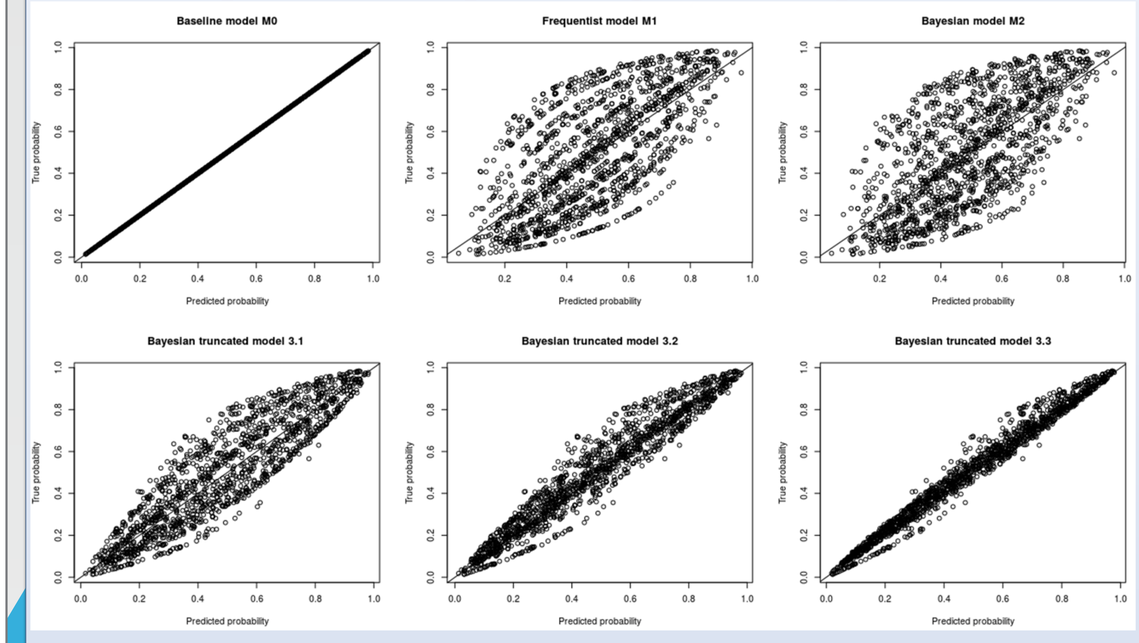
4. How to find the truncated area?

Since it's a simulation study, we know true values for the random effects.

- We first **divide the distribution** $N(0, \hat{\sigma}_u^2)$ into 2 halves, or 3 one thirds or 5 one fifths (as shown in the Figure above).
- We then simulate an expert to **choose an area from the distribution** which represents her guess of the **relative position** of new cluster *c* with regard to other clusters.
- The predicted random effect $\hat{u}_c$ for cluster *c* is subsequently **sampled from** the chosen **truncated area**.
- **Correct expert opinion** refers to a chosen truncated area where the true value of the random effect lies.

Results

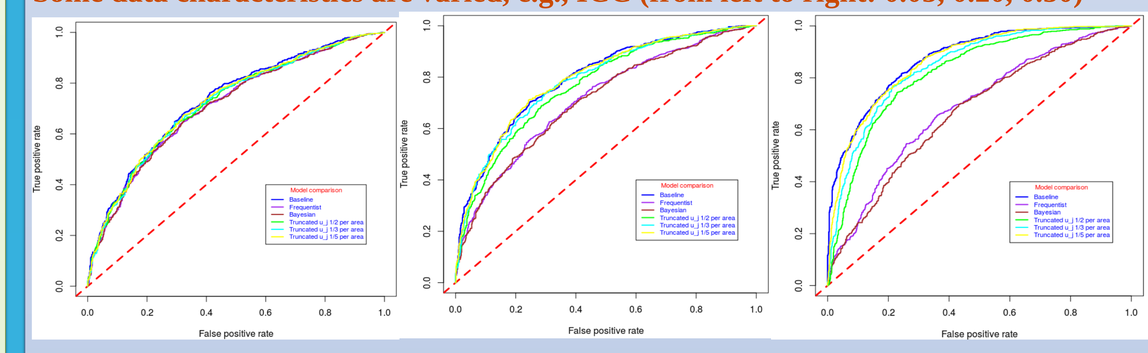
Predicted probabilities plotted against true probabilities



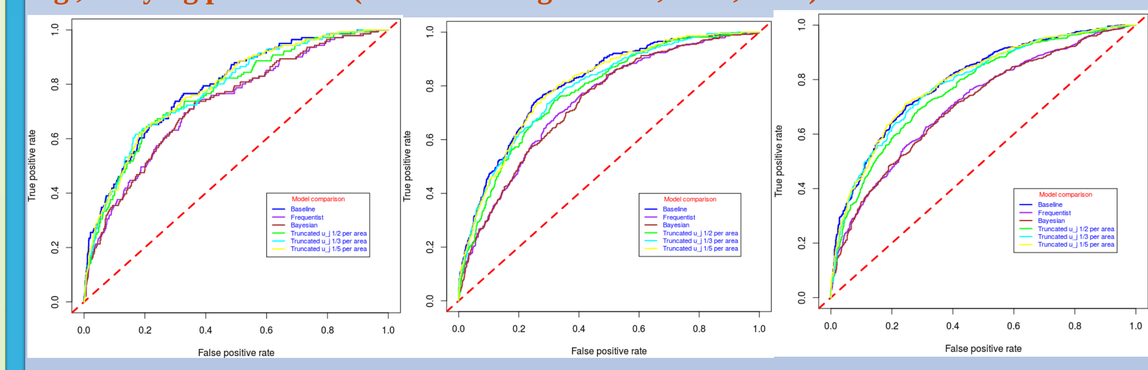
Model M0: Predicted probabilities are equivalent to the true probabilities;
Model M1: Standard model with random effects removed;
Model M2: Bayesian model with random effects attained;
Model M3.1: Bayesian model incorporating correct expert opinion sampled from half of $N(0, \hat{\sigma}_u^2)$
Model M3.2: Bayesian model incorporating correct expert opinion sampled from one third of $N(0, \hat{\sigma}_u^2)$
Model M3.3: Bayesian model incorporating correct expert opinion sampled from one fifth $N(0, \hat{\sigma}_u^2)$

	M0	M1	M2	M3.1	M3.2	M3.3
Overall Sensitivity (%)	74.3	62.2	62.4	69.6	73.5	73.0
Overall Specificity (%)	74.6	69.0	69.2	71.4	73.9	75.7
Overall accuracy (%)	74.4	65.6	65.9	70.5	73.7	74.3
Overall AUC (%)	81.9	71.8	71.6	78.3	80.9	81.7
Brier score	0.174	0.215	0.216	0.190	0.179	0.175

Some data characteristics are varied, e.g., ICC (from left to right: 0.05, 0.20, 0.50)



e.g., varying prevalence (from left to right: 10%, 25%, 50%)



Conclusion

- By using Bayesian modeling, **the random effects** for the new clusters **can be attained** rather than removed.
- The Bayesian models with informative priors for the random effects (**M3.1, M3.2 and M3.3**) **outperform** the standard frequentist model **M2** that removes the random effects.
- The **more specific the expert opinion** (i.e., smaller truncated area), the **better predictions**.

References

1. Bouwmeester et al. (2013).**DOI:** 10.1186/1471-2288-13-19.
2. Van der Drift et al. (2012). **DOI**: 10.3168/jds.2011-4417.
3.Gelman & Hill (2006). **Data Analysis using regression and multilevel / hierarchical models.** Cambridge university press.