# BAYESIAN MODELLING FOR BENEFIT-RISK BALANCE ANALYSIS: ROSIGLITAZONE FOR TYPE II DIABETES

K. Vamvourellis, K. Kalogeropoulos, L. Phillips

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

## DRUG REGULATION

Rosiglitazone is drug used, since 2000, for the treatment of Type II Diabetes. **Data that subsequently emerged about possible cardiovascular risks** resulted in the suspension of the drug in Europe and restriction in the US.

This study follows up from the work of two research groups* that did a Multiple-criteria decision analysis to assess the clinical effects of Rosiglitazone. The model focuses on 11 effects for which clinical trial data is collected.

| Favorable Effects (2) | Unfavorable Effects (9) |
|---|---|
| Binary:<br>- *Glycaemic efficacy*<br><br>Continuous:<br>- *Microvascular events* | Binary:<br>- *CHF*<br>- *CV death*<br>- *Non-CV death*<br>- *MI*<br>- *Stroke*<br>- *Macular oedema*<br>- *Bone fractures*<br>- *Bladder cancer*<br><br>Continuous:<br>- *Weight gain* |

Each subject $i$ receives a score $S$ which is a function $f_j(Y_j)$ of the measured effects $Y_j$ , weighted by their importance $w_j$

$$S = \sum_{j=1}^{11} w_j f_j(Y_j)$$

** Phillips, L.D., et al (2013). IMI Work Package 5: Report 2:b:ii Benefit - Risk Wave 2 Case Study Report: Rosiglitazone*

## OUR MODEL

The j[th] effect, if binary, is modeled as:

$$\begin{cases} Y_{ij} \sim \text{Bernoulli}(\eta_j), \\ h_j(\eta_j) = Z_{ij}, \end{cases}$$

where the link function *h* is the logit function.

Whereas continuous effects are observed directly, and are modeled as:

$$Y_{ij} = Z_{ij}, \; i = 1, \dots, N.$$

The augmented variables are modeled as a J-variate Gaussian distribution, with a covariance matrix *Σ*:

$$Z_{i:} \sim \mathcal{N}_J(\mu_J, \Sigma),$$

**The challenge here is that this parametrization of the problem does not have an identifiable likelihood†.**

*† Talhouk, A., Doucet, A. and Murphy, K. (2012). Efficient Bayesian inference for multivariate probit models with sparse inverse correlation matrices.*

## MCMC INFERENCE

We parametrize the problem in terms of the correlation matrix *R* by scaling all features to have variance *1* and use the following Gibbs algorithm:

- $\pi(Z|\mu, R, Y)$ sample each row $(y, z)$ with a Metropolis-Hastings algorithm using

$$\pi(z|y,\mu,R) \propto f(y|z)\cdot\pi(z|\mu,R)$$
$$= \prod_{j=1}^{J} \sigma(z_j)^{y_j}(1-\sigma(z_j))^{(1-y_j)} \cdot \mathcal{N}(z|\mu,R)$$

- $\pi(R|\mu, Z)$ sample *R* = *DΣD* where *W* = *ZD* and

$$\pi(\Sigma|W) = \mathcal{IW}(\Sigma; 2+N, W'W + I_J - \xi^{-1}M'M)$$

- $\pi(\mu|R, Z)$ sample using a conjugate Gaussian prior

- $\pi(\sigma_c|\mu_c, Y_c)$ variance of the continuous features is sampled using a conjugate Inverse Gamma prior

This algorithm identifies the correct values, as we confirmed with simulated experiments, but there is room for improvement in its mixing.