On parameter orthogonality and proper modelling of dispersion in PiG regression

Stephane Heritier

Monash University, Melbourne, Australia

Joint work with G. Heller (Macquarie University) and D. Couturier (Cambridge University)

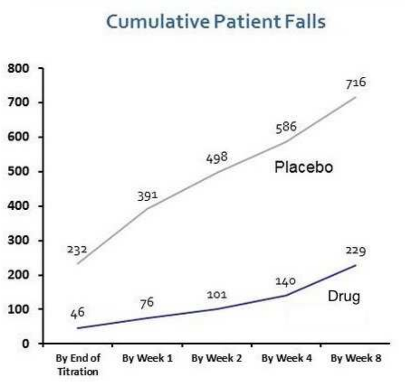
Email: stephane.heritier@monash.edu

VicBiostat seminar, 24 September 2015

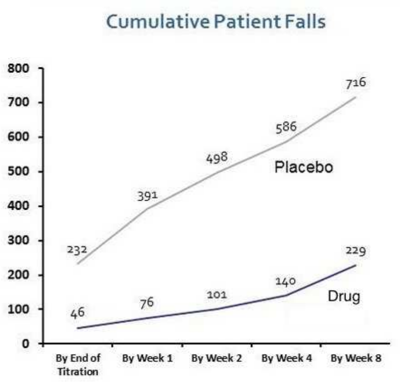
Clinical trial of drug for treatment of nOH

- Neurogenic Orthostatic Hypotension (nOH) is a sudden, dangerous fall in blood pressure when standing from a sitting or lying position.
- nOH affects patients with Parkinson's Disease (PD).
- xxxxx is a drug for controlling this condition.
- Clinical trial of xxxxx for treatment of nOH:
  - Patients randomised to receive treatment or placebo
  - $n = 197$
  - over 8 weeks
  - primary endpoint: nOH symptom score
  - secondary endpoint: self-reported number of falls

Clinical trial: results



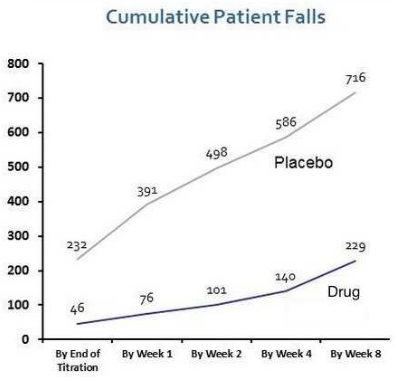
Clinical trial: results



	Treat	Control
n	105	92
Mean falls	3.4	8.7

Incidence rate ratio $= 0.39$

Clinical trial: results



	Treat	Control
n	105	92
Mean falls	3.4	8.7

Incidence rate ratio = 0.39

- Basic bootstrap 95%CI: IRR=0.39 (0.13 - 0.90)
- Fairly convincing evidence of a treatment effect

Clinical trial: results

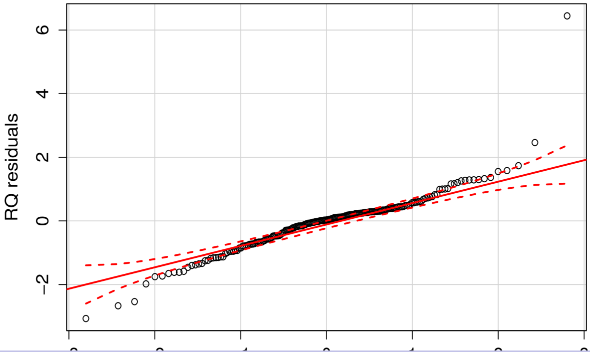
- Initial analysis of number of falls: negative binomial model
- Treatment effect not significant
- Doesn't look right

Clinical trial: results

- Initial analysis of number of falls: negative binomial model
- Treatment effect not significant
- Doesn't look right

Clinical trial: results

- Initial analysis of number of falls: negative binomial model
- Treatment effect not significant
- Doesn't look right
- NB model - residuals



Clinical trial: results (cont'd)

Clinical trial: results (cont'd)

- Looking at data again:

	Treat	Control
n	105	92
No. falls		
Mean	3.4	8.7
Variance	62.0	1388.1
Maximum	49	358

  - Treatment appears to reduce mean number of falls
  - Treatment also appears to reduce (dramatically) variance of falls
  - We need a model that reflects these features

Statistical model for number of falls

Candidate distributions for number of falls:

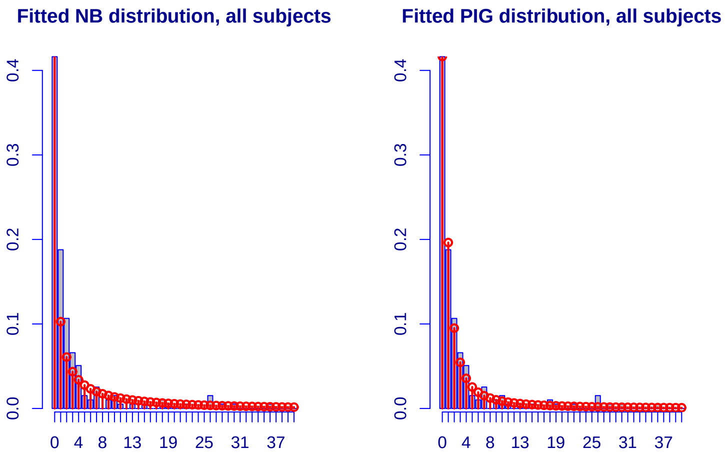
- Poisson
- compound Poisson:
  - Negative binomial
  - Poisson-inverse Gaussian (PiG)
  - Poisson-generalized inverse Gaussian (Sichel)

Statistical model for number of falls

Candidate distributions for number of falls:

- Poisson
- compound Poisson:
  - Negative binomial
  - Poisson-inverse Gaussian (PiG)
  - Poisson-generalized inverse Gaussian (Sichel)
- Zero-inflated Poisson/NB models

Statistical model for number of falls



Poisson-inverse Gaussian (PiG) distribution

$$\begin{aligned} y \,|\, \lambda &\sim \text{Poisson}(\lambda) \qquad\qquad \Rightarrow \quad y \sim \text{PiG}(\mu, \sigma) \\ \lambda &\sim \text{inverse Gaussian}(\mu, \sigma) \end{aligned}$$

Poisson-inverse Gaussian (PiG) distribution

$$y \,|\, \lambda \sim \text{Poisson}(\lambda) \qquad \Rightarrow \qquad y \sim \text{PiG}(\mu, \sigma)$$
$$\lambda \sim \text{inverse Gaussian}(\mu, \sigma)$$

$$f(y \,|\, \mu, \sigma) = \sqrt{\frac{2}{\pi\sigma}}(1 + 2\mu\sigma)^{\frac{1}{4}} e^{\frac{1}{\sigma}} \frac{\left(\mu/\sqrt{1 + 2\mu\sigma}\right)^y}{y!} K_{y-0.5}\left(\sqrt{1 + 2\mu\sigma}/\sigma\right)$$
$$y = 0, 1, 2, \ldots$$

$$E(y) = \mu$$
$$Var(y) = \mu(1 + \sigma\mu)$$

σ : dispersion parameter

Poisson-inverse Gaussian (PiG) distribution

$$\begin{aligned} y \,|\, \lambda &\sim \text{Poisson}(\lambda) \\ \lambda &\sim \text{inverse Gaussian}(\mu, \sigma) \end{aligned} \qquad \Rightarrow \qquad y \sim \text{PiG}(\mu, \sigma)$$

$$f(y \,|\, \mu, \sigma) = \sqrt{\frac{2}{\pi\sigma}} (1 + 2\mu\sigma)^{\frac{1}{4}} e^{\frac{1}{\sigma}} \frac{\left(\mu / \sqrt{1 + 2\mu\sigma}\right)^y}{y!} K_{y-0.5}\left(\sqrt{1 + 2\mu\sigma}/\sigma\right)$$

$$y = 0, 1, 2, \ldots$$

$$\begin{aligned} E(y) &= \mu \\ Var(y) &= \mu(1 + \sigma\mu) \end{aligned} \qquad \sigma : \text{ dispersion parameter}$$

- $K_\nu(x)$ is a Bessel function.
- Poisson is the limiting distribution as $\sigma \to 0$

Generalized Additive Models for Location, Scale and Shape (GAMLSS)

- Rigby and Stasinopoulos (2005) introduced Generalized Additive Models for Location, Scale and Shape (GAMLSS).
- Regression models for a wide variety of response distributions
- Modeling of mean and up to 3 shape parameters

Generalized Additive Models for Location, Scale and Shape (GAMLSS)

- Rigby and Stasinopoulos (2005) introduced Generalized Additive Models for Location, Scale and Shape (GAMLSS).
- Regression models for a wide variety of response distributions
- Modeling of mean and up to 3 shape parameters
- PiG regression:

$$y \sim \mathsf{PiG}(\mu, \sigma)$$
$$\log(\mu) = x^t \beta$$
$$\log(\sigma) = w^t \gamma$$

Statistical model for number of falls

- In the analysis of clinical trials, typically only the mean is modelled.
  - Model A: treatment effect on mean only
  - Model B: treatment effect on mean and dispersion

Statistical model for number of falls

- In the analysis of clinical trials, typically only the mean is modelled.
  - Model A: treatment effect on mean only
  - Model B: treatment effect on mean and dispersion

Statistical model for number of falls

- In the analysis of clinical trials, typically only the mean is modelled.
  - Model A: treatment effect on mean only
  - Model B: treatment effect on mean and dispersion

Model A (restricted)
$y \sim \text{PiG}(\mu, \sigma)$
$\log \mu = \beta_0 + \beta_1 x + \log t$
$\log \sigma = \gamma_0$
(similar to initial negative binomial analysis)

Model B (full)
$y \sim \text{PiG}(\mu, \sigma)$
$\log \mu = \beta_0 + \beta_1 x + \log t$
$\log \sigma = \gamma_0 + \gamma_1 x$

- x is an indicator variable for treatment
- $\log t$ is an offset term for treatment duration t.

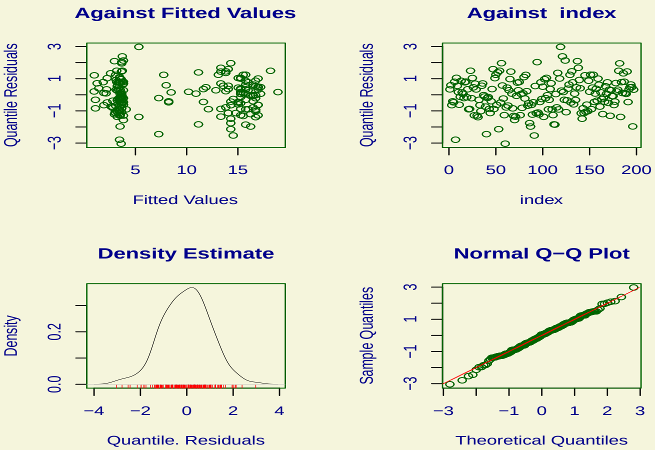
Statistical model for number of falls

	Model A (restricted)			Model B		
Parameter	estimate	s.e.	p-value	estimate	s.e.	p-value
β_0	-1.779	0.327	<0.001	-1.417	0.541	0.009
β_1	-0.322	0.337	0.341	-1.489	0.601	0.014
γ_0	2.970	0.380	<0.001	3.461	0.592	<0.001
γ_1	-	-	-	-1.667	0.706	0.002

Statistical model for number of falls

	Model A (restricted)			Model B		
Parameter	estimate	s.e.	p-value	estimate	s.e.	p-value
β_0	-1.779	0.327	$<$0.001	-1.417	0.541	0.009
β_1	-0.322	0.337	0.341	-1.489	0.601	0.014
γ_0	2.970	0.380	$<$0.001	3.461	0.592	$<$0.001
γ_1	-	-	-	-1.667	0.706	0.002

- $\hat{\beta}_1$ is sensitive to specification of the model for σ
- This is particularly bad in the clinical trials context



Against Fitted Values
Quantile Residuals
Fitted Values
Against index
Quantile Residuals
index
Density Estimate
Density
Quantile. Residuals
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles

Parameter orthogonality

The notion of parameter orthogonality means, for a two-parameter distribution $f(y \mid \mu, \theta)$:

$$E\left(\frac{\partial^2}{\partial\mu\,\partial\theta}\log f\right) = 0$$

Parameter orthogonality

The notion of parameter orthogonality means, for a two-parameter distribution $f(y \mid \mu, \theta)$:

$$E\left(\frac{\partial^2}{\partial\mu\,\partial\theta}\log f\right) = 0$$

- The MLEs $\hat{\mu}$ and $\hat{\theta}$ are asymptotically independent
- This has advantages for parameter estimation.

Parameter orthogonality

The notion of parameter orthogonality means, for a two-parameter distribution $f(y \mid \mu, \theta)$:

$$E\left(\frac{\partial^2}{\partial\mu\,\partial\theta}\log f\right) = 0$$

- The MLEs $\hat{\mu}$ and $\hat{\theta}$ are asymptotically independent
- This has advantages for parameter estimation.

Parameter orthogonality

The notion of parameter orthogonality means, for a two-parameter distribution $f(y \mid \mu, \theta)$:

$$E\left(\frac{\partial^2}{\partial\mu\,\partial\theta}\log f\right) = 0$$

- The MLEs $\hat{\mu}$ and $\hat{\theta}$ are asymptotically independent
- This has advantages for parameter estimation.
- Cox and Reid (1987), JRSSB

Parameter orthogonality

- There are several parametrizations of the PiG in the literature.
- The (μ, σ) parametrization was first proposed by Dean, Lawless, and Willmot (1989), and used by Rigby and Stasinopoulos in GAMLSS
  - appealing interpretation of σ as a Poisson overdispersion parameter
  - but μ and σ are not orthogonal

Parameter orthogonality

- There are several parametrizations of the PiG in the literature.
- The (μ, σ) parametrization was first proposed by Dean, Lawless, and Willmot (1989), and used by Rigby and Stasinopoulos in GAMLSS
  - appealing interpretation of σ as a Poisson overdispersion parameter
  - but μ and σ are not orthogonal
- Stein, Zucchini and Juritz (1987) proposed an orthogonal parametrization of the PiG:
  - Retain μ
  - Set $\alpha = \frac{\sqrt{1+2\mu\sigma}}{\sigma}$
  - μ and α are orthogonal

Orthogonal parametrization of PiG (μ, α)

$$f(y\,|\,\mu,\alpha) = \sqrt{\frac{2\alpha}{\pi}}\exp\left(\sqrt{\mu^2+\alpha^2}-\mu\right)\frac{\left(\mu\left(\sqrt{\mu^2+\alpha^2}-\mu\right)/\alpha\right)^y}{y!}K_{y-0.5}(\alpha)$$

$$E(y) = \mu$$

$$Var(y) = \mu\left(1+\frac{\mu}{\sqrt{\mu^2+\alpha^2}-\mu}\right)$$

Orthogonal parametrization of PiG (μ, α)

$$f(y \mid \mu, \alpha) = \sqrt{\frac{2\alpha}{\pi}} \exp\left(\sqrt{\mu^2 + \alpha^2} - \mu\right) \frac{\left(\mu\left(\sqrt{\mu^2 + \alpha^2} - \mu\right)/\alpha\right)^y}{y!} K_{y-0.5}(\alpha)$$

$$E(y) = \mu$$

$$Var(y) = \mu\left(1 + \frac{\mu}{\sqrt{\mu^2 + \alpha^2} - \mu}\right)$$

- $Var(y)$ has an inverse relationship with α
- Poisson is the limiting distribution as $\alpha \to \infty$

Orthogonal parametrization of PiG (μ, α)

We can specify models for μ and α :

$$y \sim \mathsf{PiG}(\mu, \alpha)$$
$$\log(\mu) = x^t \beta$$
$$\log(\alpha) = w^t \delta$$

Orthogonal parametrization of PiG (μ, α)

We can specify models for μ and α :

$$y \sim \mathsf{PiG}(\mu, \alpha)$$
$$\log(\mu) = x^t \beta$$
$$\log(\alpha) = w^t \delta$$

From orthogonality of μ and α, it follows that

$$E\left(\frac{\partial^2}{\partial \beta_j \, \partial \delta_k} \log f\right) = 0$$

i.e. the elements of β and the elements of δ are orthogonal.

Orthogonal PiG models for number of falls

Model C (restricted)
$y \sim \mathsf{PiG}(\mu, \alpha)$
$\log \mu = \beta_0 + \beta_1 x + \log t$
$\log \alpha = \delta_0$

Model D (full)
$y \sim \mathsf{PiG}(\mu, \alpha)$
$\log \mu = \beta_0 + \beta_1 x + \log t$
$\log \alpha = \delta_0 + \delta_1 x$

Orthogonal PiG models for number of falls

Model C (restricted)

$y \sim \text{PiG}(\mu, \alpha)$

$\log \mu = \beta_0 + \beta_1 x + \log t$

$\log \alpha = \delta_0$

Model D (full)

$y \sim \text{PiG}(\mu, \alpha)$

$\log \mu = \beta_0 + \beta_1 x + \log t$

$\log \alpha = \delta_0 + \delta_1 x$

	Model C			Model D		
Parameter	estimate	s.e.	p-value	estimate	s.e.	p-value
β_0	-0.865	0.632	0.171	-0.870	0.669	0.193
β_1	-2.077	0.687	0.003	-2.074	0.714	0.004
δ_0	-0.034	0.095	0.720	-0.093	0.124	0.453
δ_1	-	-	-	0.152	0.196	0.438

Orthogonal PiG models for number of falls

Model C (restricted)
$y \sim \text{PiG}(\mu, \alpha)$
$\log \mu = \beta_0 + \beta_1 x + \log t$
$\log \alpha = \delta_0$

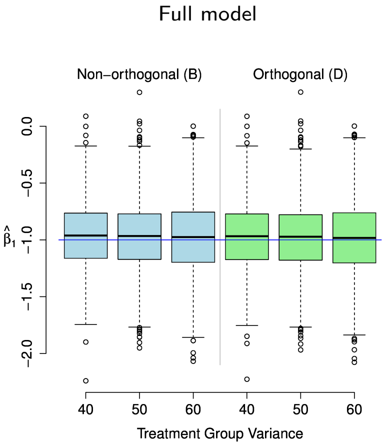
Model D (full)
$y \sim \text{PiG}(\mu, \alpha)$
$\log \mu = \beta_0 + \beta_1 x + \log t$
$\log \alpha = \delta_0 + \delta_1 x$

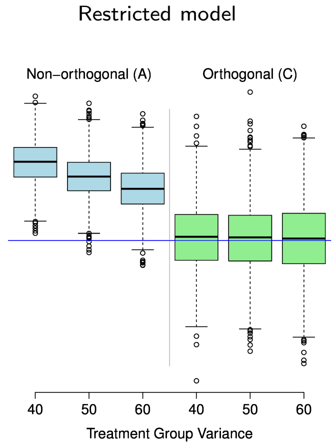
	Model C			Model D		
Parameter	estimate	s.e.	p-value	estimate	s.e.	p-value
β_0	-0.865	0.632	0.171	-0.870	0.669	0.193
β_1	-2.077	0.687	0.003	-2.074	0.714	0.004
δ_0	-0.034	0.095	0.720	-0.093	0.124	0.453
δ_1	-	-	-	0.152	0.196	0.438

- $\hat{\beta}_0$, $\hat{\beta}_1$ robust to specification of model for α
- $\hat{\beta}_1$ highly significant in both models

Simulation study 1

- control group variance = 900
- treatment group variance = 40, 50, 60
- treatment effect on mean: β_1 = -1





Simulation study 2 : Inference

- $n = 200, 500, \ldots, 1000 \quad \beta_1 = -2$
- 95% confidence intervals for β_1 (95%)
- Full (i.e. well specified) model for dispersion

Table : Coverage of 95% CI for β_1

n	gamlss	Wald Obs	Wald Asym	Sand	LRT	Bootstrap
200	89.1	89.9	89.9	81.4	96.4	86.8
500	91.8	91.9	91.7	87.5	95.9	90.3
1000	93.8	93.6	93.6	89.9	96.2	91.9

If we use the orthogonal parametrization ...

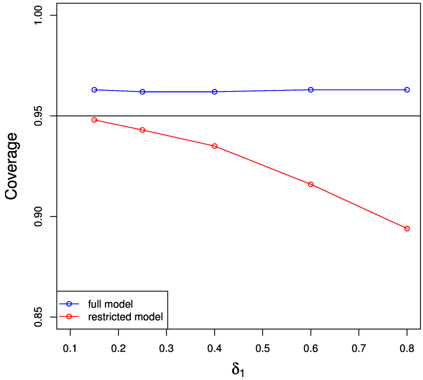
- Can we ignore the dispersion model?
- Is there a price to pay for not modelling the dispersion?

Simulation study 2 (cont'd)

- $n = 200 \quad \beta_1 = -2$
- Penalised likelihood ratio confidence intervals for β_1 (95%)

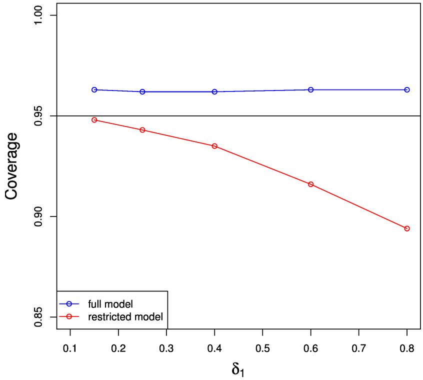
Simulation study 2 (cont'd)

- $n = 200 \quad \beta_1 = -2$
- Penalised likelihood ratio confidence intervals for β_1 (95%)



Simulation study 2 (cont'd)

- $n = 200 \quad \beta_1 = -2$
- Penalised likelihood ratio confidence intervals for β_1 (95%)



- Falls data: $\hat{\delta}_1 \simeq 0.15$

Conclusions

- When modelling mean and dispersion, we need to consider parametrization of the response distribution.
  - In exponential family, the mean μ and exponential dispersion parameter ϕ are orthogonal (so GLMs are OK).
  - Outside exponential family .. beware of non-orthogonal parametrization

Conclusions

- When modelling mean and dispersion, we need to consider parametrization of the response distribution.
  - In exponential family, the mean μ and exponential dispersion parameter ϕ are orthogonal (so GLMs are OK).
  - Outside exponential family .. beware of non-orthogonal parametrization
- RCTs: what exactly do we mean by "treatment effect"?
  - treatment effect on the mean only
  - treatment effect on the mean and dispersion

Conclusions

- When modelling mean and dispersion, we need to consider parametrization of the response distribution.
  - In exponential family, the mean μ and exponential dispersion parameter ϕ are orthogonal (so GLMs are OK).
  - Outside exponential family .. beware of non-orthogonal parametrization
- RCTs: what exactly do we mean by "treatment effect"?
  - treatment effect on the mean only
  - treatment effect on the mean and dispersion
- Inference : LRT 95% CI better (may requires proper modelling of dispersion)

References

- Cox, D. R. and N. Reid (1987). Parameter orthogonality and approximate conditional inference. Journal of the Royal Statistical Society. Series B, 49(1), 1–39.
- Dean, C., J. Lawless, and G. Willmot (1989). A mixed Poisson–inverse-Gaussian regression model. *Canadian Journal of Statistics* 17 (2), 171–181.
- Rigby, R. and D. Stasinopoulos (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54 (3), 507– 554.
- Stein, G. Z., W. Zucchini and J. M. Juritz (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association* 82 (399), 938–944.