Dynamic Analysis of Recurrent Event Data, with Application to Infant Diarrhoea in Brazil

Ørnulf Borgan

Rosemeire Fiaccone

and

Robin.Henderson@ncl.ac.uk


ANCISCO
CONDE
CANDEIAS
MADRE DE DEUS
SIMÕES FILHO
BAÍA DE TODOS OS SANTOS
ITAPARICA
SALVADOR
VERA CRUZ
BRASIL
BAHIA
SALVADOR
AMÉRICA DO SUL
OCEANO PACÍFICO
OCEANO ATLÂNTICO

Blue Bay Project

Public works and education in the areas of sanitation and environment executed by the Bahia State Government, Brazil

$>$ \$1 billion

Belgica 1996

Belgica 2002

Data

Daily data from household survey

Home visits over 455 days October 2000 to January 2002

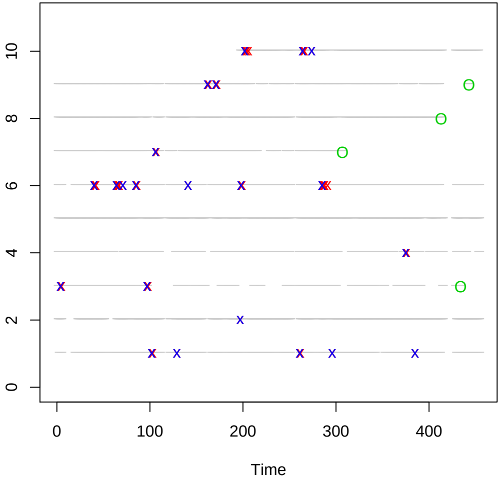
We use 926 children with $>$ 90 days follow-up

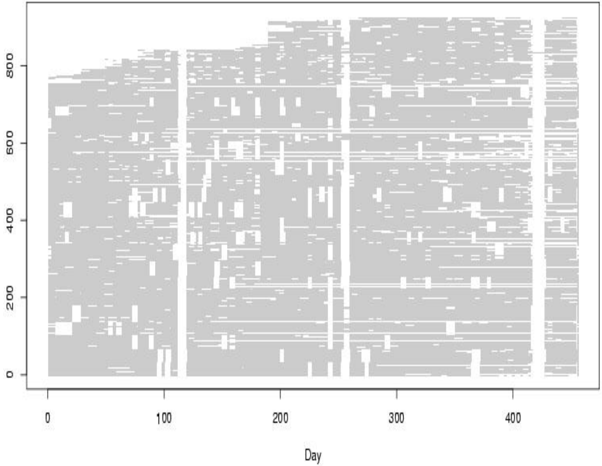
Age $<$ 3 years on entry

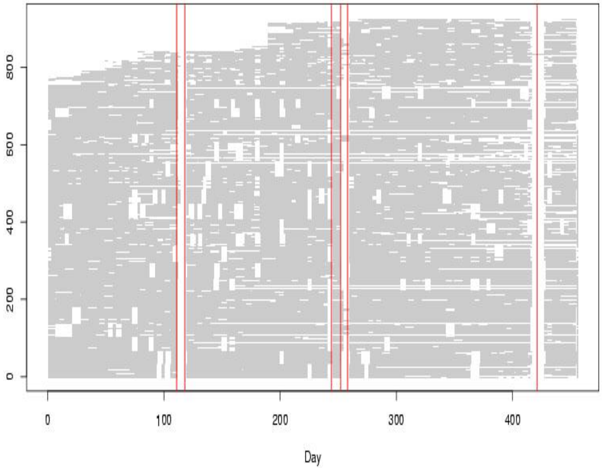
Various social, demographic and economic characteristics collected at the beginning of the study

Interested in *incidence* **and** *prevalence* **of** *diarrhoea* **over calendar time**

Point Processes




0
200
400
600
800
0
100
200
300
400
Day


0
200
400
600
800
0
100
200
300
400
Day


CAFE
SANTA
CLARA
NOVA SCHIN
TIM

Counting Processes & Additive Model

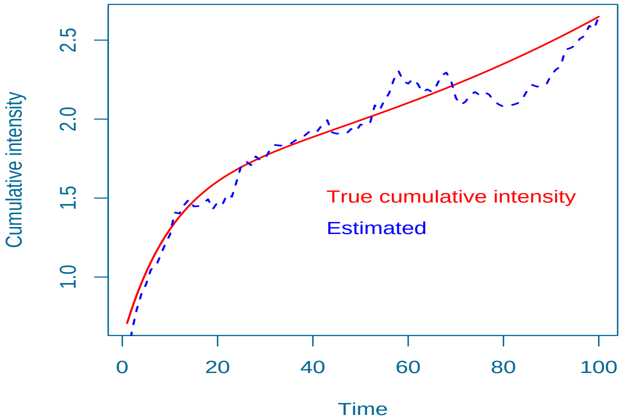
- $\mathbf{N_i(t) = \Lambda_i(t) + M_i(t)}$

Counting Processes & Additive Model

- $\mathbf{N_i(t) = \Lambda_i(t) + M_i(t)}$
- $\mathbf{\Lambda_i(t) = \int_0^t Y_i(u)\alpha_i(u)du = \int_0^t \lambda_i(u)du}$

Counting Processes & Additive Model

- $\mathbf{N_i(t) = \Lambda_i(t) + M_i(t)}$
- $\mathbf{\Lambda_i(t) = \int_0^t Y_i(u)\alpha_i(u)du = \int_0^t \lambda_i(u)du}$
- $\mathbf{\lambda_i(t) = Y_i(t)\{\beta_0(t) + x_{i1}(t)\beta_1(t) + x_{i2}(t)\beta_2(t) + \ldots\}}$


Cumulative intensity
1.0
1.5
2.0
2.5
0
20
40
60
80
100
Time
True cumulative intensity
Estimated

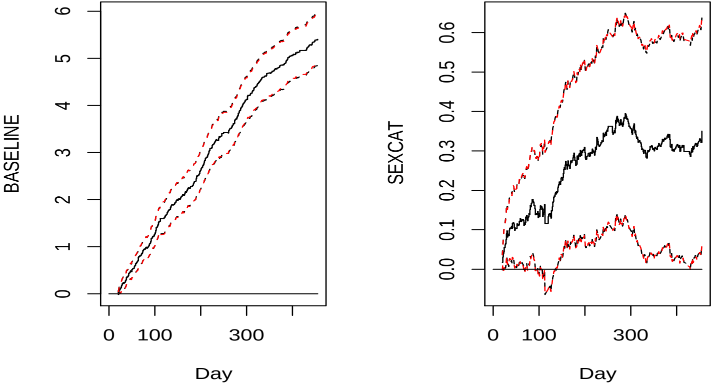
- $\mathbf{B_j(t)} = \int_{\mathbf{0}}^{\mathbf{t}} \beta_{\mathbf{j}}\mathbf{(u)du}$

- $\mathbf{B_j(t) = \int_0^t \beta_j(u)du}$
- $\mathbf{dN(t) = X(t)dB(t) + dM(t)}$

- $\mathbf{B_j(t) = \int_0^t \beta_j(u)du}$
- $\mathbf{dN(t) = X(t)dB(t) + dM(t)}$
- $\mathbf{\hat{B}(t) = \int_0^t \left(X^T(u)X(u)\right)^{-1} X^T(u)dN(u) = \int_0^t H(u)dN(u)}$

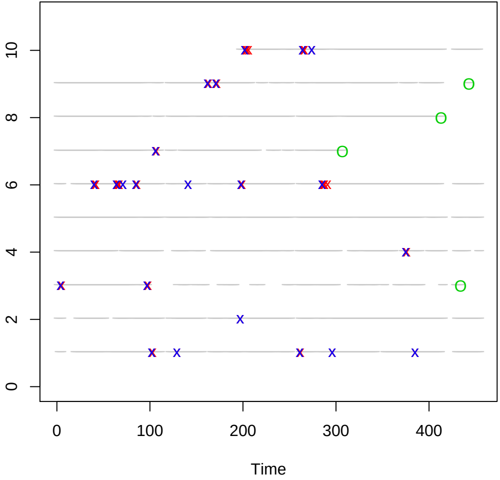
- $\mathbf{B_j(t) = \int_0^t \beta_j(u)du}$
- $\mathbf{dN(t) = X(t)dB(t) + dM(t)}$
- $\mathbf{\hat{B}(t) = \int_0^t \left(X^T(u)X(u)\right)^{-1} X^T(u)dN(u) = \int_0^t H(u)dN(u)}$
- $\mathbf{\hat{B}(t) - B(t) = \int_0^t H(u)dM(u)}$

- $B_j(t) = \int_0^t \beta_j(u)du$
- $dN(t) = X(t)dB(t) + dM(t)$
- $\hat{B}(t) = \int_0^t \left(X^T(u)X(u)\right)^{-1} X^T(u)dN(u) = \int_0^t H(u)dN(u)$
- $\hat{B}(t) - B(t) = \int_0^t H(u)dM(u)$
- Standard results apply for correctly specified models, ie $E[dN(t)|\mathcal{F}_{t-}]$

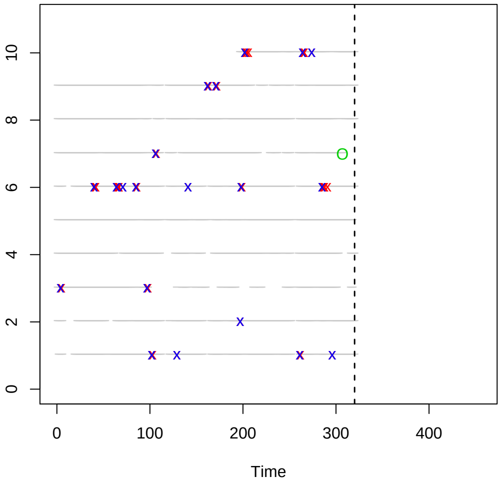


BASELINE
0
1
2
3
4
5
6
0
100
300
Day
SEXCAT
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
100
300
Day

- $\mathbf{B_j(t) = \int_0^t \beta_j(u)du}$
- $\mathbf{dN(t) = X(t)dB(t) + dM(t)}$
- $\mathbf{\hat{B}(t) = \int_0^t \left(X^T(u)X(u)\right)^{-1} X^T(u)dN(u) = \int_0^t H(u)dN(u)}$
- $\mathbf{\hat{B}(t) - B(t) = \int_0^t H(u)dM(u)}$
- **Standard results apply for correctly specified models, ie $\mathbf{E[dN(t)|\mathcal{F}_{t-}]}$**

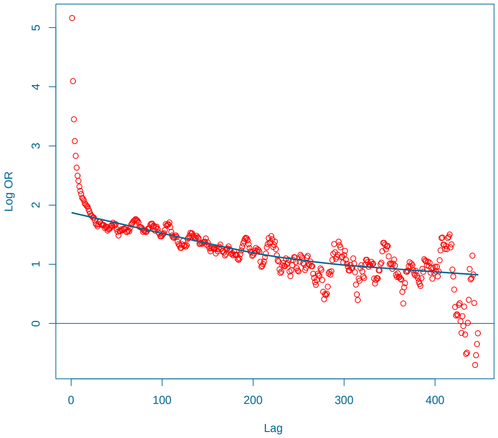


0
2
4
6
8
10
0
100
200
300
400
Time


0
2
4
6
8
10
0
100
200
300
400
Time

Lorelogram (Heagerty and Zeger, 1998)

(log odds ratio for event or no events separated by lags 1,2,3,...)



Dynamic Covariates (Aalen et al 2004, Fosen et al 2005)

Dynamic covariates $Z(t)$ are functions of individual-specific histories

Eg $Z(t)$=previous diarrhoea rate (episodes/time)

Dynamic Covariates (Aalen et al 2004, Fosen et al 2005)

Dynamic covariates $Z(t)$ are functions of individual-specific histories

Eg $Z(t)$=previous diarrhoea rate (episodes/time)

Example: test for effect of rain-affected accommodation

Model	Rain-affected	Previous episode rate
No dynamic	3.70	

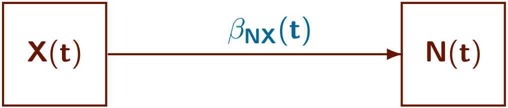
Dynamic Covariates (Aalen et al 2004, Fosen et al 2005)

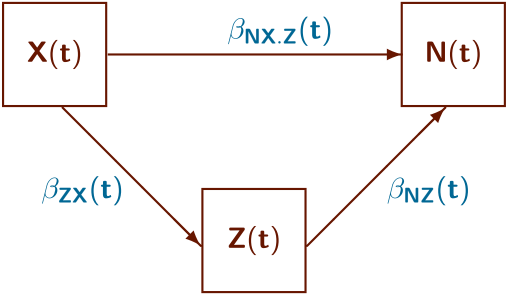
Dynamic covariates $Z(t)$ are functions of individual-specific histories

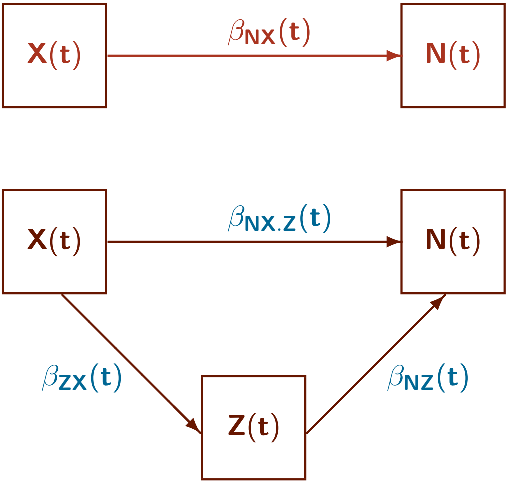
Eg $Z(t)$=previous diarrhoea rate (episodes/time)

Example: test for effect of rain-affected accommodation

Model	Rain-affected	Previous episode rate
No dynamic	3.70	
Include $Z(t)$	1.53	6.78


X(t)
β_NX(t)
N(t)


X(t)
βNX.Z(t)
N(t)
βZX(t)
Z(t)
βNZ(t)



$$\beta_{NX}(t) = \beta_{NX.Z}(t) + \beta_{NZ}(t)\beta_{ZX}(t)$$

Solution

- **Assume**

$$\mathbf{Z(t)} = \mathbf{X(t)}\gamma\mathbf{(t)} + \mathbf{R(t)}$$

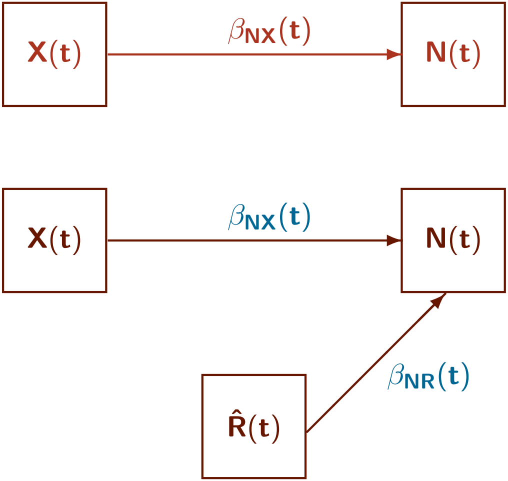
Solution

- **Assume**

$$\mathbf{Z(t)} = \mathbf{X(t)}\gamma\mathbf{(t)} + \mathbf{R(t)}$$

- **Use**

$$\hat{\mathbf{R}}\mathbf{(t)} = \mathbf{Z(t)} - \left(\mathbf{X^T(t)X(t)}\right)^{-1}\mathbf{X^T(t)Z(t)}$$


X(t)
βNX(t)
N(t)
X(t)
βNX(t)
N(t)
R̂(t)
βNR(t)

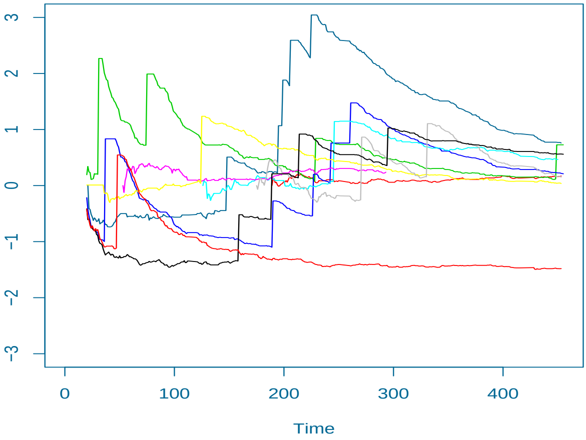
Example: test for effect of rain-affected accommodation

$Z(t)$=previous diarrhoea rate (episodes/time)

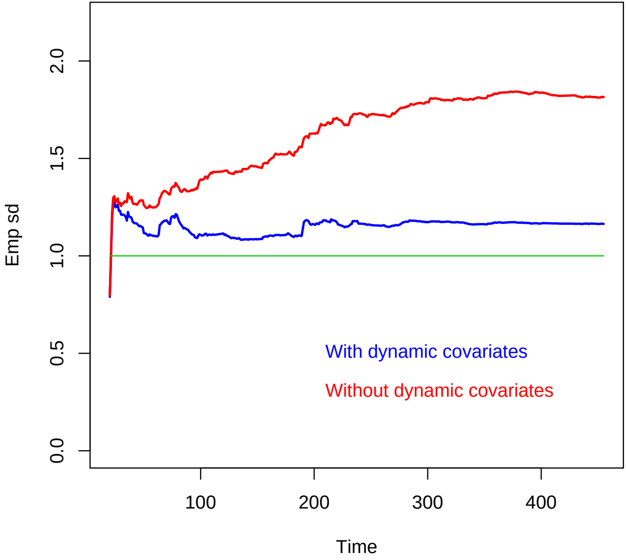
Model	Rain-affected	Previous episode rate
No dynamic	3.70	
Include $Z(t)$	1.53	6.78
Include $R(t)$	3.79	6.77

Residual processes: $\mathbf{dN(t) - X(t)d\hat{B}(t)}$

Standardised by model SD



Empirical SD of Standardised Residuals



Another Diagnostic (Diggle et al 2007)

- $\mathbf{Cov}(\mathbf{M}(\mathbf{t_0}), \mathbf{M}(\mathbf{t})) = \mathbf{Var}(\mathbf{M}(\mathbf{t_0})) \qquad \mathbf{t} > \mathbf{t_0}$

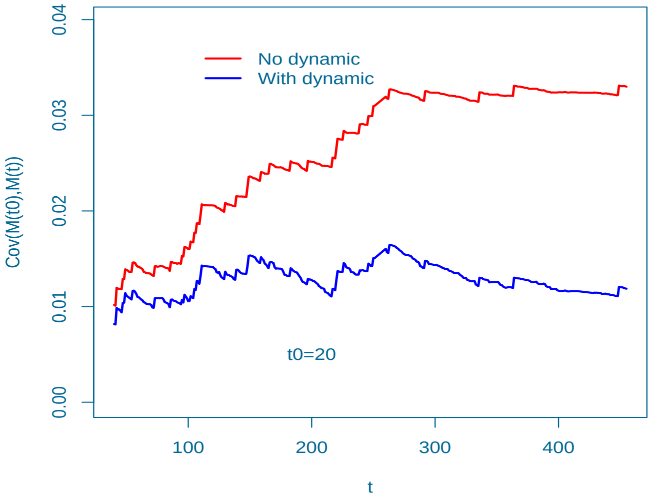
Another Diagnostic (Diggle et al 2007)

- $\mathbf{Cov}(\mathbf{M}(\mathbf{t_0}), \mathbf{M}(\mathbf{t})) = \mathbf{Var}(\mathbf{M}(\mathbf{t_0})) \qquad \mathbf{t} > \mathbf{t_0}$
- **Plot of** $\mathbf{Cov}(\hat{\mathbf{M}}(\mathbf{t_0}), \hat{\mathbf{M}}(\mathbf{t}))$ **against t should be flat**

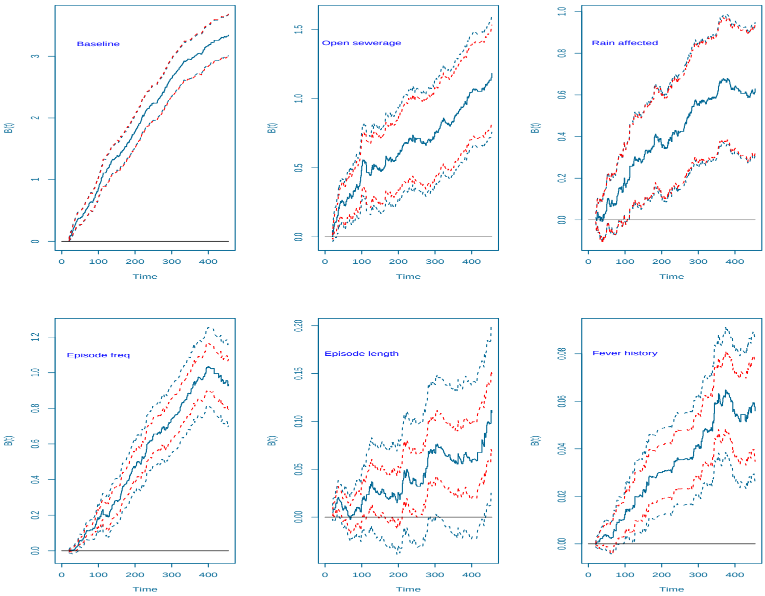
Another Diagnostic (Diggle et al 2007)

- $\mathbf{Cov}(\mathbf{M}(\mathbf{t_0}), \mathbf{M}(\mathbf{t})) = \mathbf{Var}(\mathbf{M}(\mathbf{t_0})) \qquad \mathbf{t} > \mathbf{t_0}$
- **Plot of** $\mathbf{Cov}(\hat{\mathbf{M}}(\mathbf{t_0}), \hat{\mathbf{M}}(\mathbf{t}))$ **against** $\mathbf{t}$ **should be flat**
- **Could try various** $\mathbf{t_0}$

Martingale Covariance Plot

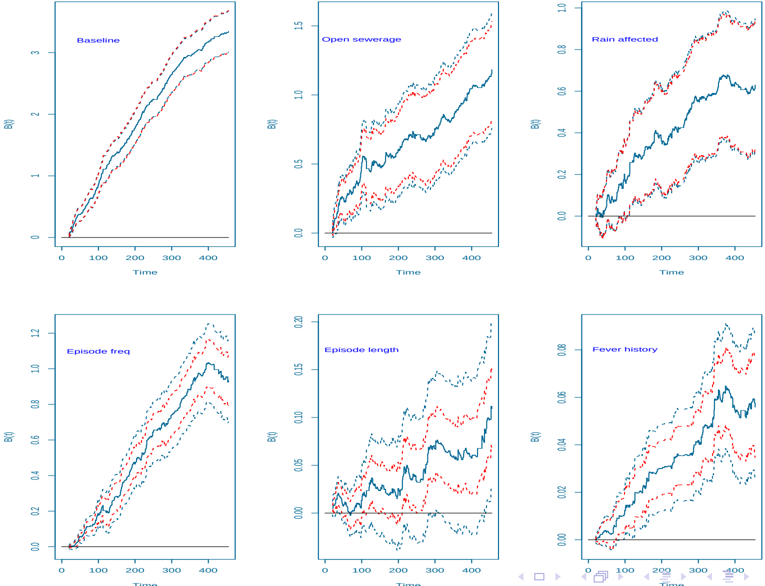


Some Incidence Results



Some Incidence Results

Also crowding, mother's age, child age, male



Frailty

- **True model:** $\mathbf{dN(t) = Z \times X(t)dB(t) + dM_Z(t)}$

Frailty

- **True model:** $\mathbf{dN(t) = Z \times X(t)dB(t) + dM_Z(t)}$
- **Fitted model:** $\mathbf{dN(t) = E[Z|\mathcal{F}_{t-}]X(t)dB(t) + dM(t)}$

Frailty

- **True model:** $dN(t) = Z \times X(t)dB(t) + dM_Z(t)$
- **Fitted model:** $dN(t) = E[Z|\mathcal{F}_{t-}]X(t)dB(t) + dM(t)$
- $Z_i \sim \Gamma(1/\xi, 1/\xi)$
- $$E[Z_i|\mathcal{F}_{t-}] = \frac{1 + \xi N_i(t^-)}{1 + \xi \Lambda_i(t^-)}$$

Frailty

- **True model:** $\mathbf{dN(t) = Z \times X(t)dB(t) + dM_Z(t)}$
- **Fitted model:** $\mathbf{dN(t) = E[Z|\mathcal{F}_{t-}]X(t)dB(t) + dM(t)}$
- $\mathbf{Z_i \sim \Gamma(1/\xi, 1/\xi)}$
-
$$\mathbf{E[Z_i|\mathcal{F}_{t-}] = \frac{1 + \xi N_i(t^-)}{1 + \xi \Lambda_i(t^-)}}$$
- $\mathbf{d\hat{B}_\xi(t) = \left(X^T(t)X(t)\right)^{-1} X^T(t)dN(t)/E[Z|\mathcal{F}_{t-}]}$

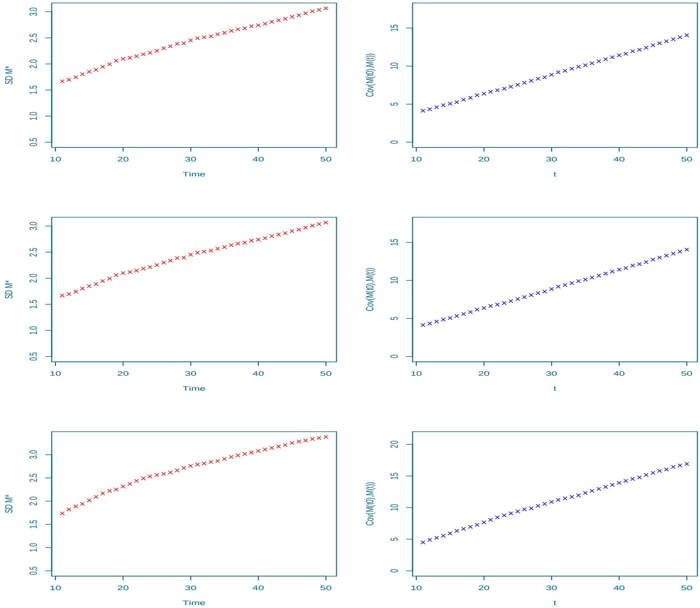
Frailty

- True model: $dN(t) = Z \times X(t)dB(t) + dM_Z(t)$
- Fitted model: $dN(t) = E[Z|\mathcal{F}_{t-}]X(t)dB(t) + dM(t)$
- $Z_i \sim \Gamma(1/\xi, 1/\xi)$
- $$E[Z_i|\mathcal{F}_{t-}] = \frac{1 + \xi N_i(t^-)}{1 + \xi \Lambda_i(t^-)}$$
- $d\hat{B}_\xi(t) = \left(X^T(t)X(t)\right)^{-1} X^T(t)dN(t)/E[Z|\mathcal{F}_{t-}]$
- Estimate ξ by maximum (negative binomial) likelihood for $N(\tau)$

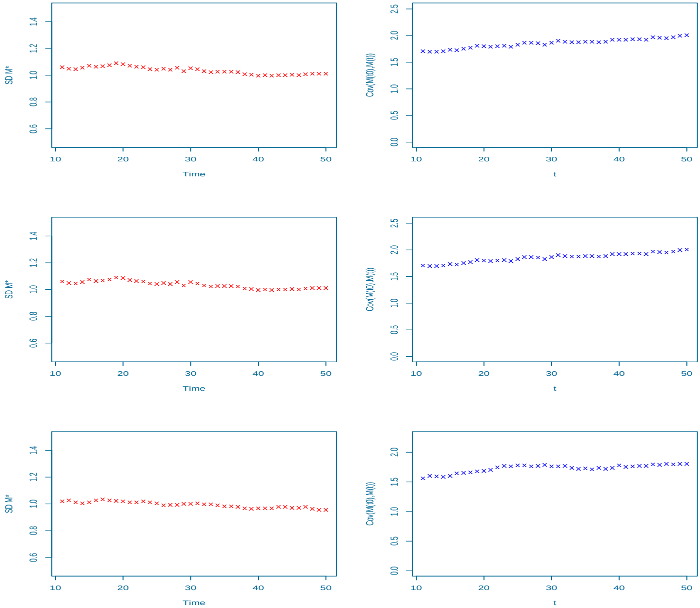
Simulations

- $\tau = 50$, discrete time
- Two binary covariates
- $\beta_0(t) = 0.1 \quad \beta_1(t) = \beta_2(t) = 0.05$
- Fit
  1. Fixed effects only
  2. Dynamic covariate model A: $D(t) = N(t)/t$
  3. Dynamic covariate model B: $D(t) = \{N(t) - N(t-15)\}/15$
  4. Dynamic covariate model C: $D(t) = \{N(t) - N(t-10)\}/10$
  5. Frailty model

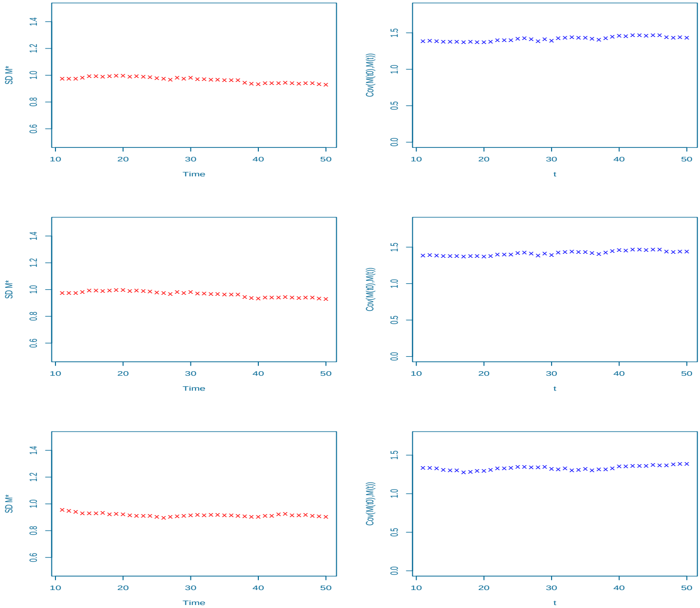
Diagnostics - fixed effects fit, frailty data



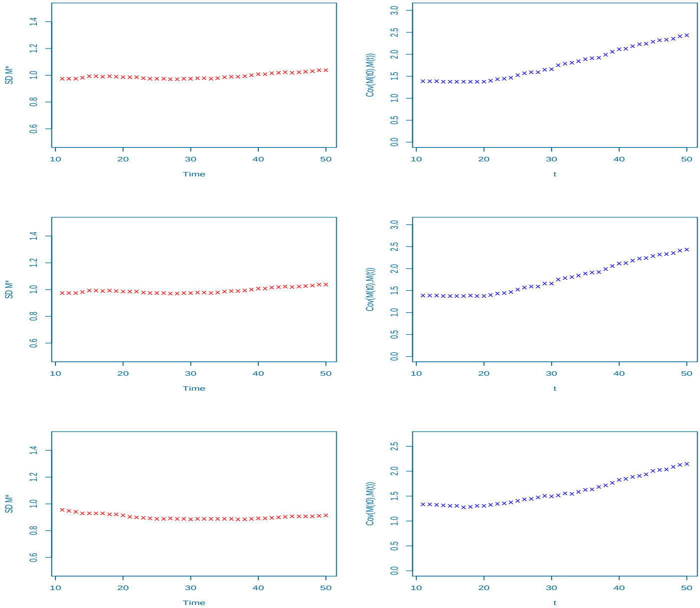
Diagnostics - frailty fit, frailty data



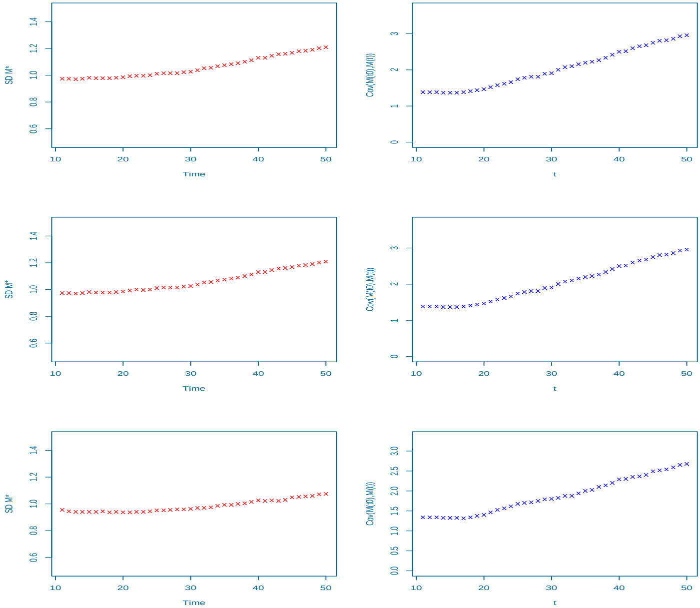
Diagnostics - dynamic fit A, frailty data



Diagnostics - dynamic fit B, frailty data

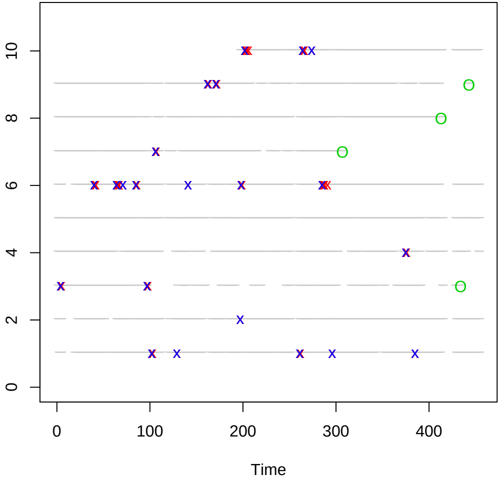


Diagnostics - dynamic fit C, frailty data



Sample of Blue Bay Data

Diarrhoea episodes in 926 children followed for 455 days



Multivariate gamma frailty (Henderson & Shimakura 2003)

- **Time varying $\mathbf{Z_i(t)}$ $\mathbf{t = 1, 2, \ldots \tau}$**

Multivariate gamma frailty (Henderson & Shimakura 2003)

- **Time varying $\mathbf{Z_i(t)}$ $\mathbf{t} = \mathbf{1}, \mathbf{2}, \ldots \tau$**
- **Defined by Laplace transform**

$$\mathbf{E}\{\exp(-\mathbf{u}'\mathbf{Z})\} = |\mathbf{I} + \xi\mathbf{C} \times \mathrm{diag}(\mathbf{u})|^{-\mathbf{1}/\xi}$$

- $\mathbf{C_{jk}} = \rho^{|\mathbf{j}-\mathbf{k}|/\mathbf{2}}$

Multivariate gamma frailty (Henderson & Shimakura 2003)

- **Time varying $\mathbf{Z_i(t)}$ $\mathbf{t} = \mathbf{1}, \mathbf{2}, \ldots \tau$**
- **Defined by Laplace transform**

$$\mathbf{E}\{\exp(-\mathbf{u}'\mathbf{Z})\} = |\mathbf{I} + \xi\mathbf{C} \times \mathrm{diag}(\mathbf{u})|^{-\mathbf{1}/\xi}$$

- $\mathbf{C_{jk}} = \rho^{|\mathbf{j}-\mathbf{k}|/\mathbf{2}}$
- $\mathbf{Z_i(t)} \sim \mathbf{\Gamma}(\mathbf{1}/\xi, \mathbf{1}/\xi) \quad \mathbf{Corr}(\mathbf{Z_i(t)}, \mathbf{Z_i(u)}) = \rho^{|\mathbf{t}-\mathbf{u}|}$

Multivariate gamma frailty (Henderson & Shimakura 2003)

- **Time varying $\mathbf{Z_i(t)}$ $\mathbf{t} = \mathbf{1}, \mathbf{2}, \dots \tau$**
- **Defined by Laplace transform**

$$\mathbf{E}\{\exp(-\mathbf{u}'\mathbf{Z})\} = |\mathbf{I} + \xi\mathbf{C} \times \mathrm{diag}(\mathbf{u})|^{-\mathbf{1}/\xi}$$

- $\mathbf{C_{jk}} = \rho^{|\mathbf{j}-\mathbf{k}|/\mathbf{2}}$
- $\mathbf{Z_i(t)} \sim \mathbf{\Gamma}(\mathbf{1}/\xi, \mathbf{1}/\xi) \quad \mathbf{Corr}(\mathbf{Z_i(t)}, \mathbf{Z_i(u)}) = \rho^{|\mathbf{t}-\mathbf{u}|}$
- $\mathbf{dN(t)} = \mathbf{Z(t)X(t)dB(t)} + \mathbf{dM_Z(t)}$

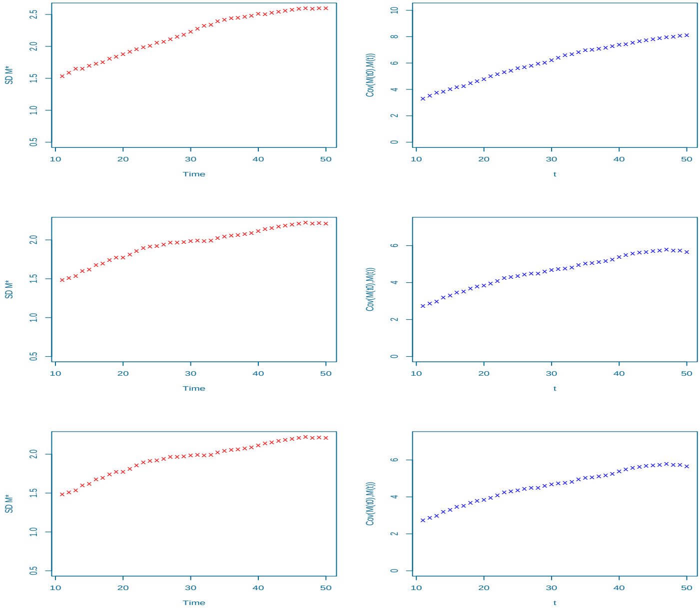
Multivariate gamma frailty (Henderson & Shimakura 2003)

- **Time varying $\mathbf{Z_i(t)}$ $\mathbf{t = 1, 2, \ldots \tau}$**
- **Defined by Laplace transform**

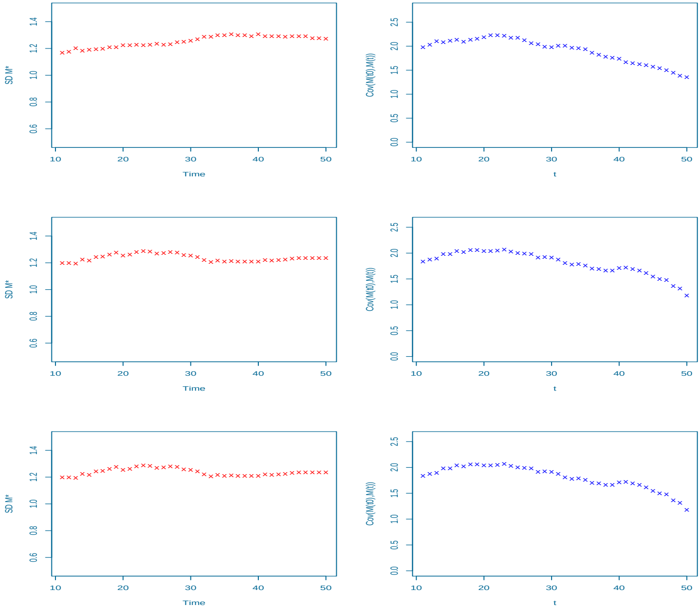
$$\mathbf{E}\{\exp(-\mathbf{u}'\mathbf{Z})\} = |\mathbf{I} + \xi\mathbf{C} \times \mathrm{diag}(\mathbf{u})|^{-\mathbf{1}/\xi}$$

- $\mathbf{C_{jk}} = \rho^{|\mathbf{j}-\mathbf{k}|/\mathbf{2}}$
- $\mathbf{Z_i(t)} \sim \mathbf{\Gamma}(\mathbf{1}/\xi, \mathbf{1}/\xi)$ $\quad$ $\mathbf{Corr(Z_i(t), Z_i(u))} = \rho^{|\mathbf{t}-\mathbf{u}|}$
- $\mathbf{dN(t) = Z(t)X(t)dB(t) + dM_Z(t)}$
- **Sims have $\rho = \mathbf{0.95}$, correlation 0.6 at lag 10**

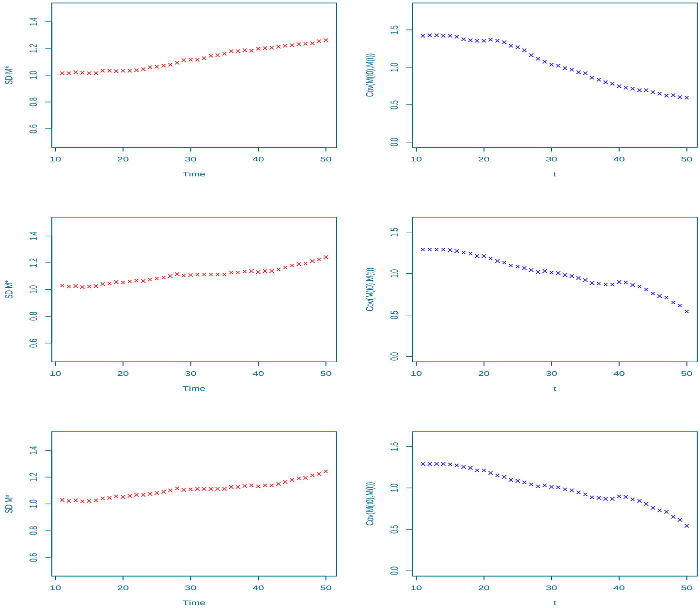
Diagnostics - fixed effects fit, correlated frailty data



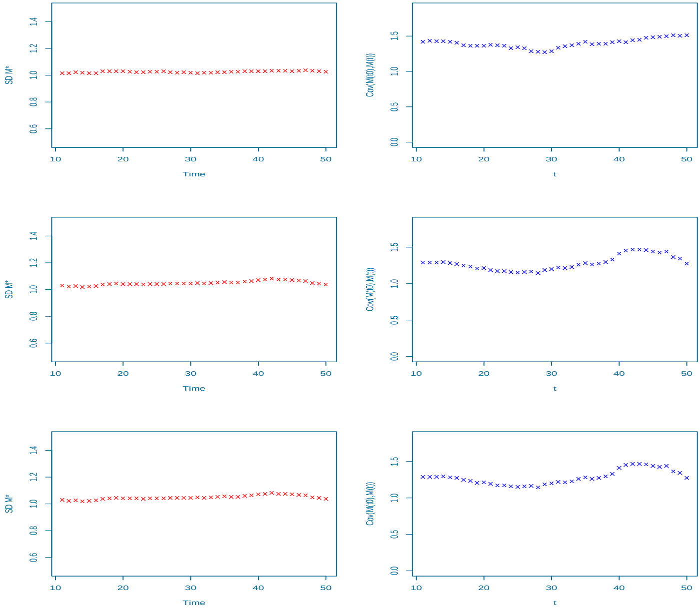
Diagnostics - frailty fit, correlated frailty data



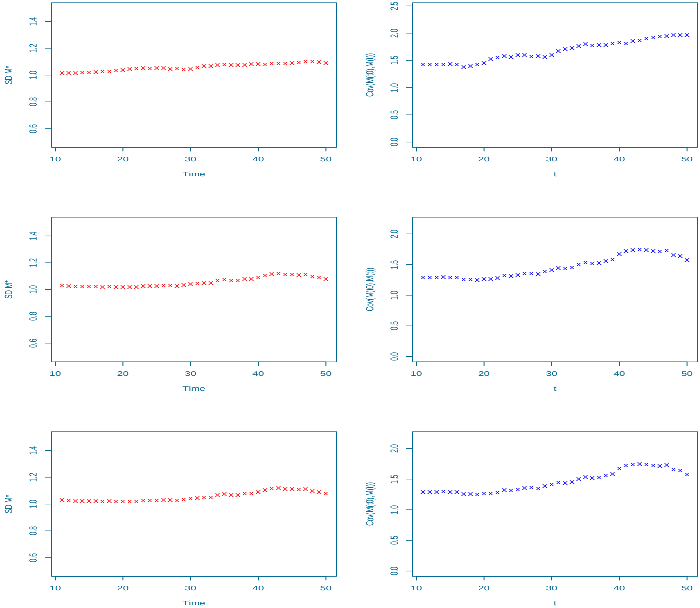
Diagnostics - dynamic fit A, correlated frailty data



Diagnostics - dynamic fit B, correlated frailty data



Diagnostics - dynamic fit C, correlated frailty data



How to fit?