



LEIDS UNIVERSITAIR MEDISCH CENTRUM

Analyzing Longitudinal Data With A Floor Level At Zero

Saskia le Cessie^{1,2}, Friedo W. Dekker¹, Diana C. Grootendorst¹

¹Department of Clinical Epidemiology

²Department of Medical Statistics and Bioinformatics

Leiden University Medical Centre



Dutch follow-up study of renal patients

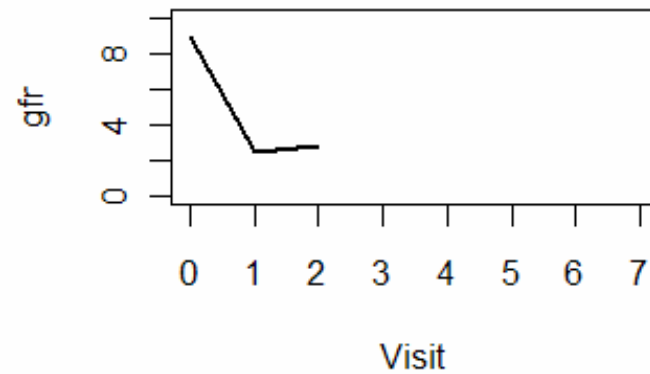
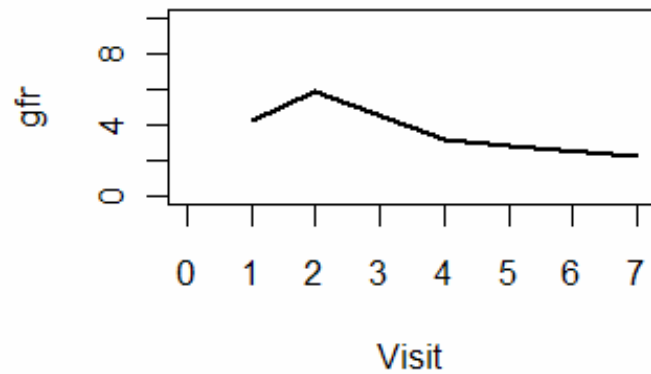
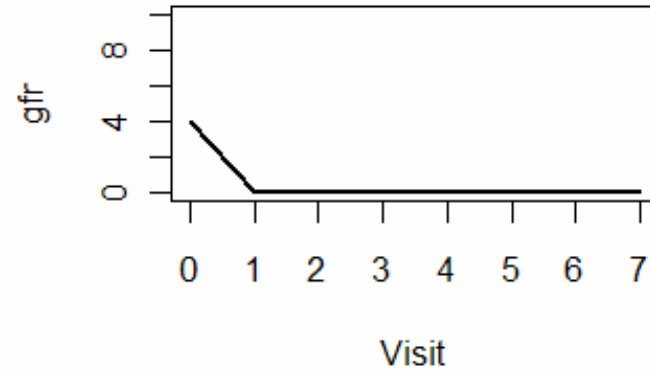
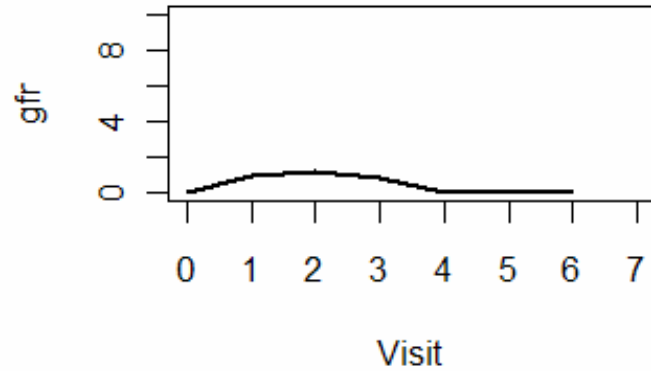
- 1526 end-stage renal failure patients who start dialysis
- Measurements at start dialysis, 3 months, 6 months and thereafter every 6 months (Here follow-up of 3 years)
- Two different forms of dialysis : hemodialysis (HD) and peritoneal dialysis(PD)
- Outcome: renal function (Glomerular filtration rate, GFR)
- Goal: model the pattern of GFR over time

Some problems

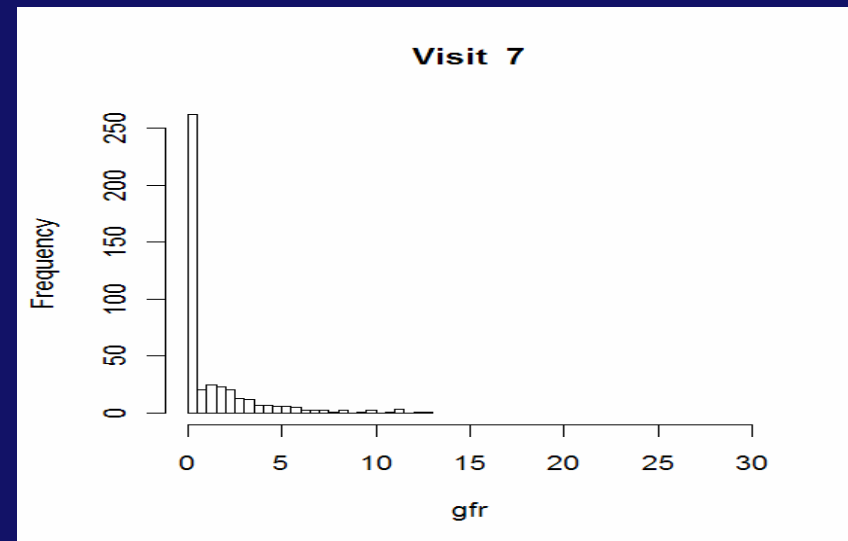
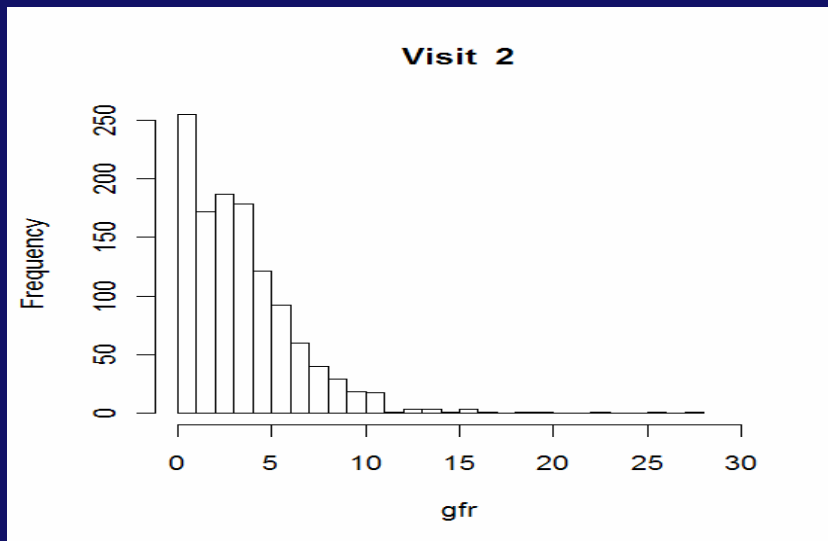
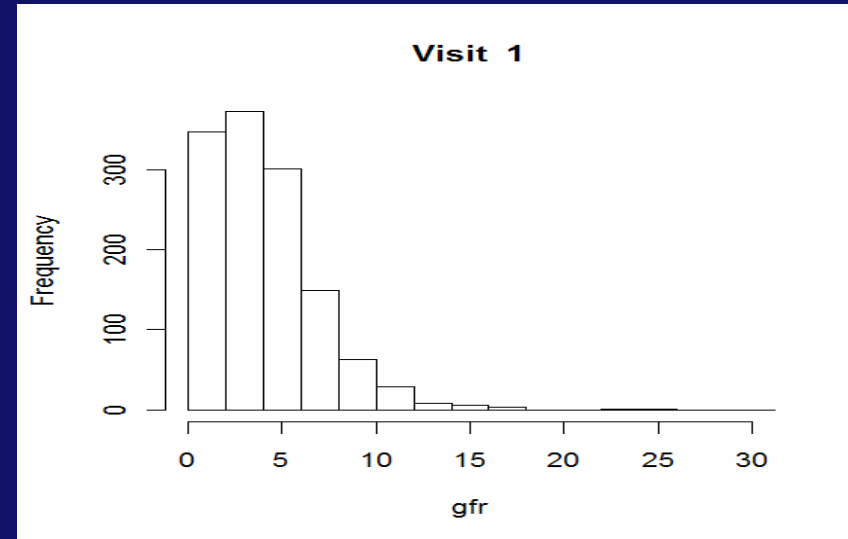
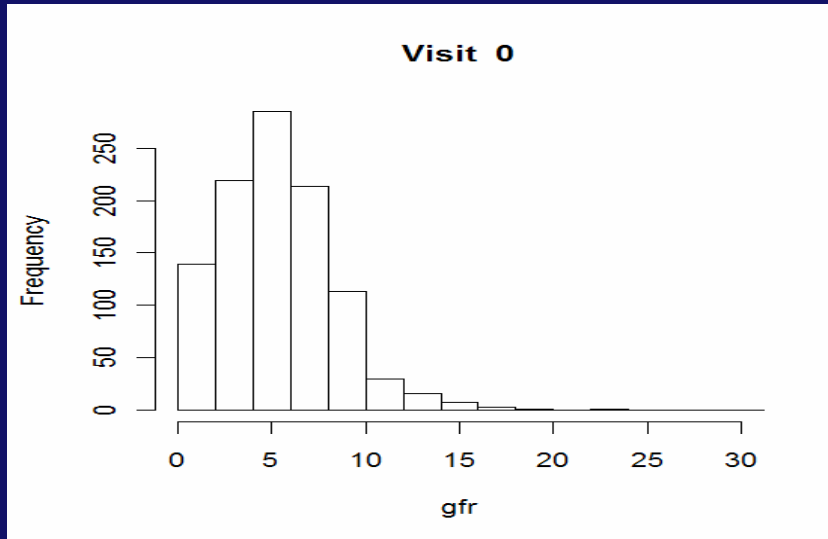
- The kidneys can stop working completely. Then GFR is per definition equal to 0. This is called anuria.
- Sometimes there are incidental GFR=0 measurements
- Patients with PD have on average a larger GFR value at start of dialysis.

GFR patterns for 4 different patients

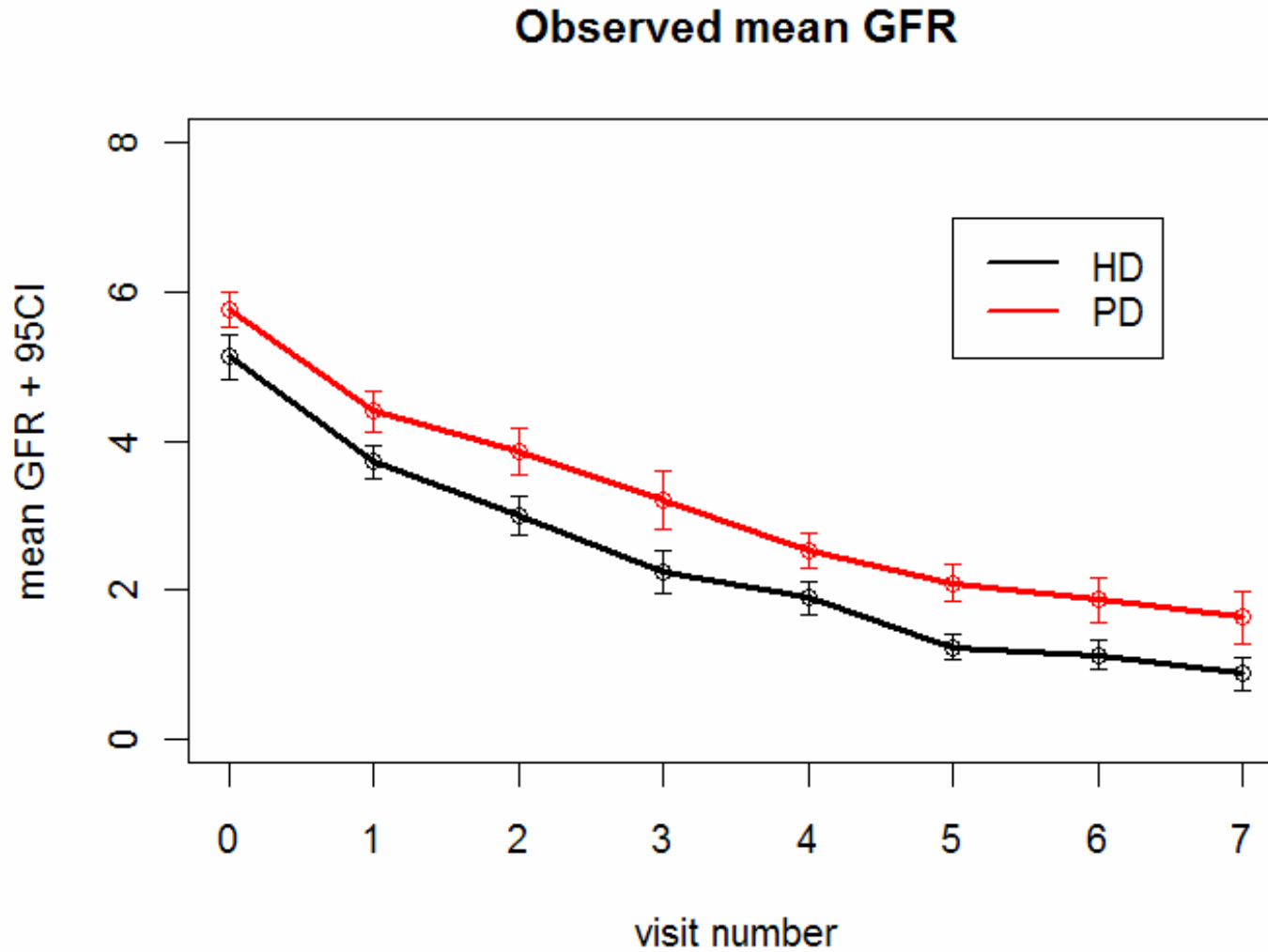
Examples of gfr patterns over time



The distribution of the gfr measurements at different time points



Observed means over time



Question: how to model this type of data?

- Three approaches
 - Linear mixed models
 - Two-part mixed models
 - Transition models (Markov approach)

Approach 1: linear mixed model

- Remove all measurements where patients are anuric (two subsequent GFR=0 measurements).
- Then perform a standard repeated measures analysis.
- Underlying idea: model the GFR trajectory before anuria

What are you doing here?

- You do not model the mean GFR over time. This model implicitly imputes values for those patients with GFR =0
- Observations are left out if GFR=0. Not missing not at random

Second approach: Two-part mixed models (Tooze 2002)

- Assume that GFR measurements (notation Y_{ij}) result from a mixture distribution:

- $f(y) = \begin{cases} \Pr(Y = 0) & \text{if } y = 0 \\ (1 - \Pr(Y = 0))g(y) & \text{if } y > 0 \\ 0 & \text{if } y < 0 \end{cases}$

- Occurrence variable $R_{ij} = \begin{cases} 0 & \text{if } Y_{ij} = 0 \\ 1 & \text{if } Y_{ij} > 0 \end{cases}$

- Model the occurrence $\Pr(R_{ij}=1)$ and the distribution of $Y_{ij}|Y_{ij}>0$ (the intensity)

Model:

- Occurrence by logistic regression with random person effect U_i
- $\text{logit}(\text{Pr}(R_{ij}=1)) = \theta X_{ij} + U_i$
- Intensity with a linear model with random person effect V_i
 - $\log(Y_{ij}) | Y_{ij} > 0 \sim N(\beta X_{ij} + V_i, \sigma_e^2)$
- Random effects are jointly normal

$$\begin{bmatrix} U_i \\ V_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix} \right)$$

- Likelihood is rather complex
- Tooze(2002) wrote a macro using PROC NLMIXED in SAS to maximize the likelihood

	Occurence	Intensity
Intercept	7.71 (0.54)	1.24 (0.05)
therapy	1.09 (0.24)	0.20 (0.04)
visit1	-3.77 (0.39)	-0.31 (0.02)
visit2	-5.07 (0.39)	-0.46 (0.02)
visit3	-6.42 (0.41)	-0.67 (0.02)
visit4	-7.62 (0.43)	-0.77 (0.03)
visit5	-8.61 (0.45)	-0.93 (0.03)
visit6	-9.25 (0.47)	-1.01 (0.03)
visit7	-10.06 (0.49)	-1.09 (0.04)
Variance random effect	13.18(1.17)	0.39 (0.02)
Covariance random effects	1.46 (0.10)	

	Occurrence	Intensity	
Intercept	7.71 (0.54)	1.24 (0.05)	<ul style="list-style-type: none"> • PD patients have higher intensity and lower probability on GFR=0 measurement • There were no significant therapy*visit interactions
therapy	1.09 (0.24)	0.20 (0.04)	
visit1	-3.77 (0.39)	-0.31 (0.02)	
visit2	-5.07 (0.39)	-0.46 (0.02)	
visit3	-6.42 (0.41)	-0.67 (0.02)	
visit4	-7.62 (0.43)	-0.77 (0.03)	
visit5	-8.61 (0.45)	-0.93 (0.03)	
visit6	-9.25 (0.47)	-1.01 (0.03)	
visit7	-10.06 (0.49)	-1.09 (0.04)	
Variance random effect	13.18(1.17)	0.39 (0.02)	
Covariance random effects	1.46 (0.10)		

- Correlation between repeated measures is rather simple (random intercept model)
- If random effects are correlated: estimation takes very long
However, ignoring the correlation could yield biased estimates(Su, Tom, Farewell (2009))
- It does not use the fact that a patient is anuric after two GFR=0 measurements

Approach 3. Transition models (Markov approach)

- Use transition models (Diggle, Liang, Zeger (1994))

- Rewrite multivariate density

$$f(y_{i1}, y_{i2}, \dots, y_{iJ}) = f(y_{i1}) f(y_{i2} | y_{i1}) \dots f(y_{iJ} | y_{iJ-1}, \dots, y_{i1})$$

Markov assumption:

- $f(y_{ij} | y_{ij-1}, \dots, y_{i1}) = f(y_{ij} | y_{ij-1})$

- Likelihood $\prod_i f(y_{i1}) \prod_{i,j>1} f(y_{ij} | y_{ij-1})$

- Maximize both parts separately

Model for $f(y_{ij}/y_{ij-1})$

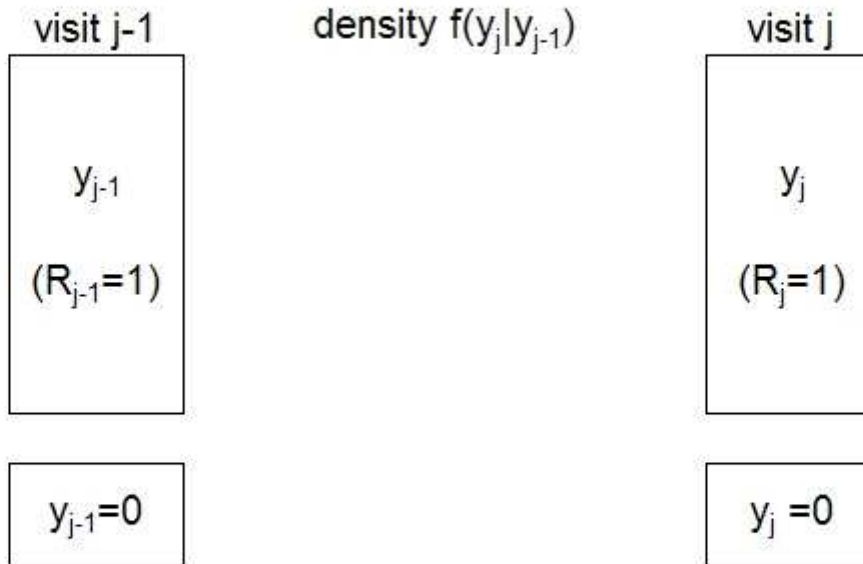
- Use two part mixture distribution:

$$f(y_{ij} | y_{ij-1}) = \begin{cases} \Pr(R_{ij} = 0 | y_{ij-1}) & \text{if } y_{ij} = 0, y_{ij-1} > 0 \\ \Pr(R_{ij} = 1 | y_{ij-1})g(y_{ij} | y_{ij-1}) & \text{if } y_{ij} > 0, y_{ij-1} > 0 \\ 1 & \text{if } y_{ij} = 0, y_{ij-1} = 0 \\ 0 & \text{if } y_{ij} > 0, y_{ij-1} = 0 \end{cases}$$

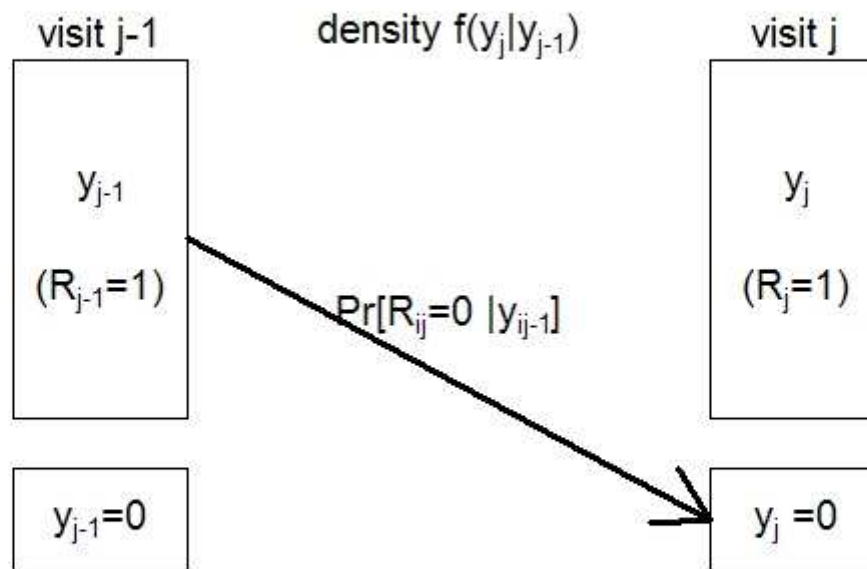
- with the occurrence variable

$$R_{ij} = \begin{cases} 0 & \text{if } Y_{ij} = 0 \\ 1 & \text{if } Y_{ij} > 0 \end{cases}$$

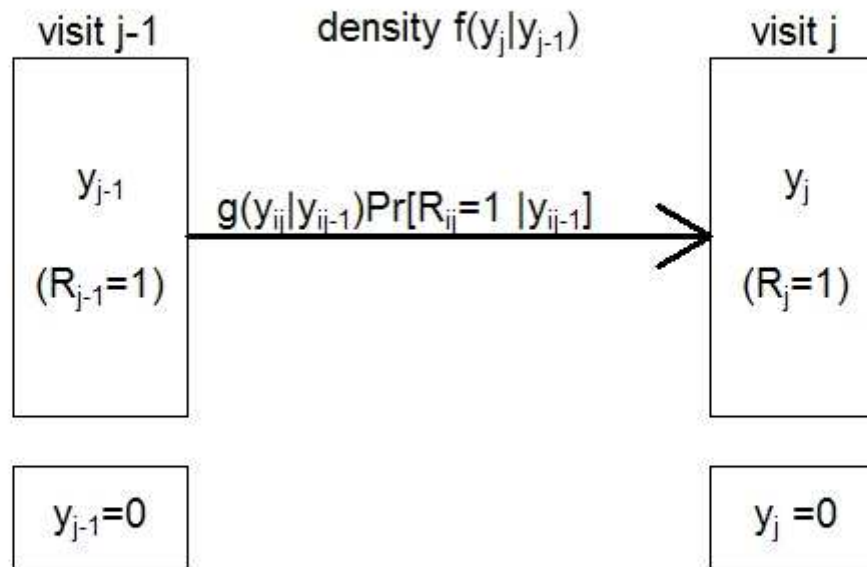
Model for $f(y_{ij}/y_{ij-1})$



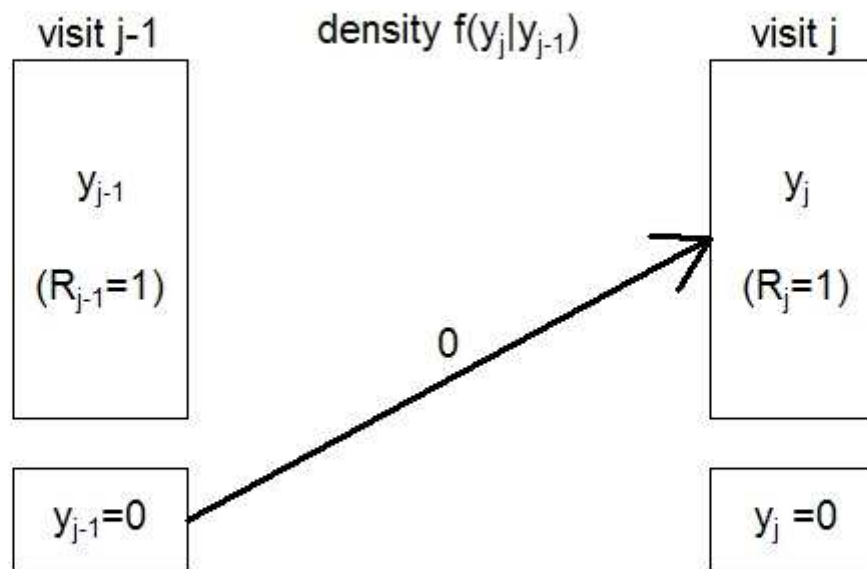
Model for $f(y_{ij}/y_{ij-1})$



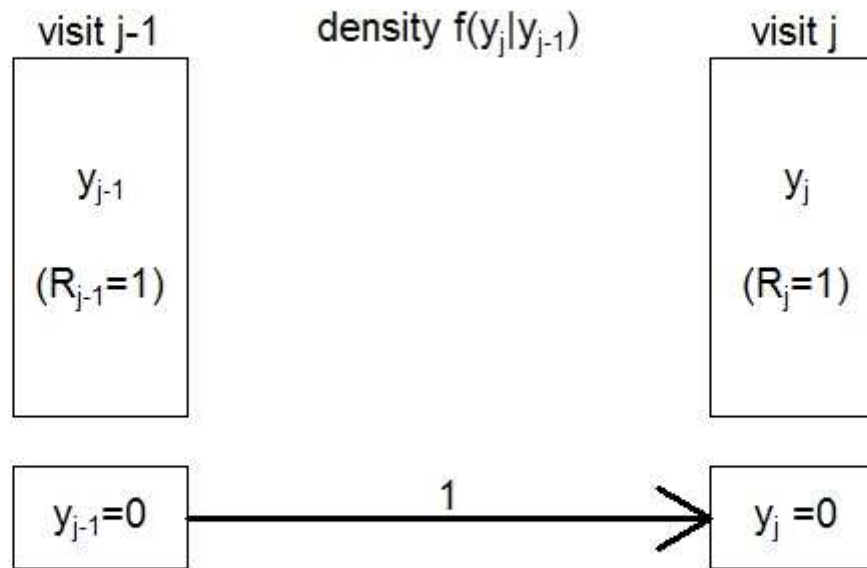
Model for $f(y_{ij}/y_{ij-1})$



Model for $f(y_{ij}/y_{ij-1})$



Model for $f(y_{ij}/y_{ij-1})$



Applying this model

- Restructure data
- Logistic model for $\Pr(R_{ij}=1|y_{ij-1})$
 - $\text{logit}(\Pr(R_{ij}=1|y_{ij-1})) = \theta_1 X_{ij} + \theta_2 y_{ij-1}$
- Linear regression model for $g(y_{ij-1}|y_{ij-1})$ (or for $\log(y_{ij-1})$)
 - $g(y_{ij-1}|y_{ij-1}) = N(\beta_1 X_{ij} + \beta_2 y_{ij-1}, \sigma_e^2)$

Application to our dataset

- We distinguished first zero and second (subsequent) zero.
- Two logistic models:
 - Probability on first zero ($R_{ij}=0$, $R_{ij-1}=1$)
 - Probability on second zero ($R_{ij}=0$, $R_{ij-1}=0$)
- After second zero, a patient is aneuric (all subsequent measures are 0)
- Two intensities:
 - Y_{ij} given $R_{ij}=1$ and $R_{ij-1}=1$
 - Y_{ij} given $R_{ij}=1$ and $R_{ij-1}=0$

Results, occurrence of first zero

- $\text{logit}(R_{ij}=1) = 0.12 \text{ (se } 0.16) + 0.30 \text{ (se } 0.10) * \text{therapy} + 1.33 * \log(\text{GFR}_{ij-1})$
- PD patients have a $\exp(0.30) = 1.35$ higher odds of $R=1$ ($\text{GFR} > 0$) given the previous GFR value

No significant effects of: $\text{GFR}_{ij-1} * \text{therapy}$, GFR_{ij-2} , , visit number

Occurrence of anuria (second zero)

- $\Pr (R_{ij}=1 | R_{ij-1}=0) = 0.50 + 0.15 (\text{se } 0.12)^* \text{ therapy} - 1.17 \log(\text{GFR}_{ij-2})$
- No significant difference between two treatment groups

Intensity when $GFR_{ij} > 0$ and $GFR_{ij-1} > 0$

- Markov assumption did not hold; Current GFR depends on previous two measurements
- Significant interaction between visit and effect of previous GFR
 - Effects of GFR_{j-1} on visit 1 was clearly different
- Visit 1 : $\log(GFR_{ij}) = 0.21 + 0.05 \text{ (se } 0.04) * \text{therapy} + 0.61 * \log(GFR_{ij-1})$
- Visit 2 to 7 : $\log(GFR_{ij}) = -0.19 + 0.04 \text{ (se } 0.02) * \text{therapy} + 0.73 * \log(GFR_{ij-1}) + 0.18 \log(GFR_{ij-2})$.
- PD patients have on average $\exp(0.04) = 1.04$ times higher GFR, given previous GFR responses ($p=0.04$)

Intensity when $GFR_{ij} > 0$ and $GFR_{ij-1} = 0$

- $\log(GFR_{ij}) = -0.50 - 0.15 \text{ (se } 0.12) * \text{therapy} + 1.17 * \log(GFR_{ij-2})$

Conclusion from transition model

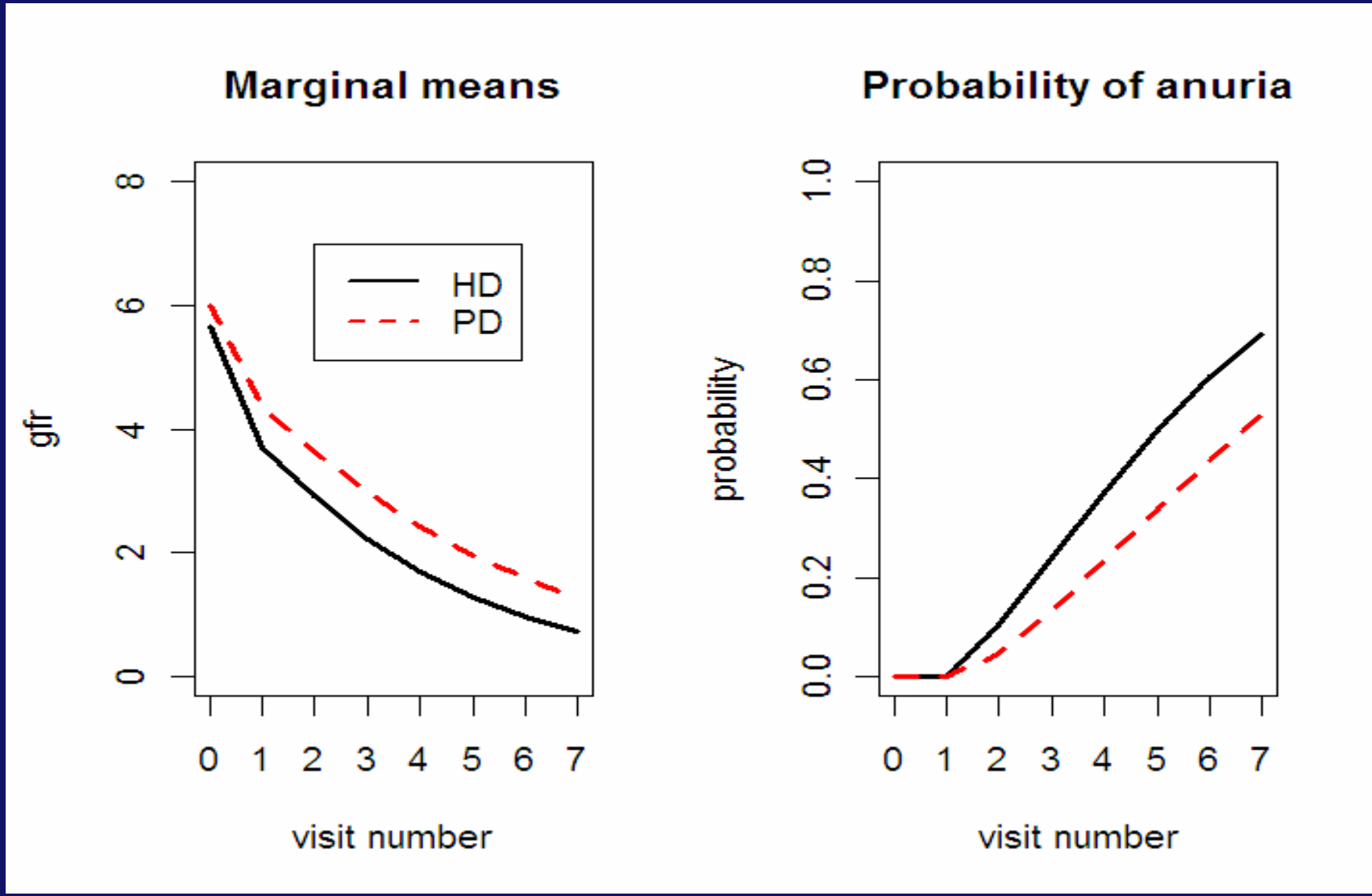
- PD patients have a higher occurrence probability and a have slightly higher GFR, after correction for previous GFR measurements
- PD patients have a higher mean intensity, given the previous GFR measurement

Marginal means

- $E[Y_j]$ can be estimated
 - Analytically, using recursively that $E[Y_j] = E [E[Y_j|Y_{j-1}]]$
 - By simulation from joint distribution

$$f(y_{i1}) f(y_{i2}|y_{i1}) \dots f(y_{iJ}|y_{iJ-1}, \dots, y_{i1})$$

Marginal means



- Transition models using a two part mixture distribution can be used to model data with zero's
- Advantages of this approach:
 - Can be performed with standard software
 - Yields interpretable parameters
 - Corrects for baseline differences