

Bayesian Data Analysis Workshop  
Espoo, May 6-7, 2013

Modelling and analysis of time  
in-homogeneous recurrent  
event processes  
in a heterogeneous population:  
A case study  
of HRTs

Elja Arjas (UH, THL, UiO)

Based on joint work with Madhuchhanda Bhattacharjee  
The work was sponsored by, and reported to, Schering in 2005.

# Background

- For menopausal women, hormone replacement therapy (HRT) provides substantial long term health benefits, in addition to relief of menopausal symptoms.
- Despite these potential benefits, compliance with HRT is reportedly low, mainly because of irregular or withdrawal bleeding.
- Therefore, in order to be able to predict the likely course and outcome of an HRT, it is essential to study the associated bleeding pattern.

# Background (cont'd)

- The bleeding is expected to completely cease after some duration (which can be dependent on the type of the therapy), resulting in an expected transience in the process.
- However this process has been observed to be non-monotone, highly time in-homogeneous and also individual dependent, i.e. heterogeneous across the study population.

# Motivation

- Here we propose a model for the statistical analysis of the bleeding patterns of different HRTs, as captured by the bleeding diaries maintained by individual patients.
- If the proposed method is able to capture the main features of the alternative treatments, as well as provide reasonable predictions on how individual patients are likely to respond, this information can be useful in clinical practice.
- It could be viewed as a form of "personalized medicine" in the context.

# Main types of HRT

- scHRT: Sequential administration of oestrogen and progestogen, usually producing a regular and therefore predictable bleeding pattern (two different types of scHRT considered here).
- ccHRT: Continuous administration of these hormones, producing a much less predictable bleeding pattern.

# Description of data

- Bleeding information is usually collected in the form of diaries, each diary recording bleeding incidents during 90 consecutive days.
- Here we consider a data set (collected and provided to us by Leiras) containing information on three different HRTs, of which two were ccHRT and one scHRT.
- Data on 163 subjects, for approx. one year each.
- Of these, 54 received type-1 ccHRT, 56 received type-2 ccHRT, and the remaining 53 were given scHRT.

# Description of data (cont'd)

- Bleeding can be of different magnitudes.
- The common practice has been to dichotomize the level into two categories, called “Bleeding” and “Spotting”.
- Here each day in a diary was coded as “B” if the patient experienced bleeding, “S” if she experienced spotting, and “N” otherwise.
- Most subjects spent more than 80 percent of their time in state “N”.

# Three 'randomly chosen' individuals ...

Table 1: No. and percentage of B, S and N days for three subjects

Subject no.	Total N days	Total S days	Total B days	Total duration
A	648	58	14	706
B	650	70	0	720
C	664	54	2	718

Subject no.	percentage N days	percentage S days	percentage B days	No. of episodes
A	90.0	8.1	1.9	69
B	90.3	9.7	0.0	36
C	92.2	7.5	0.3	9

# Three 'randomly chosen' individuals (cont'd)

Figure 1: Graphical presentation of the N (=1), S(=2) and B(=3) episodes in three selected individuals with comparable summary

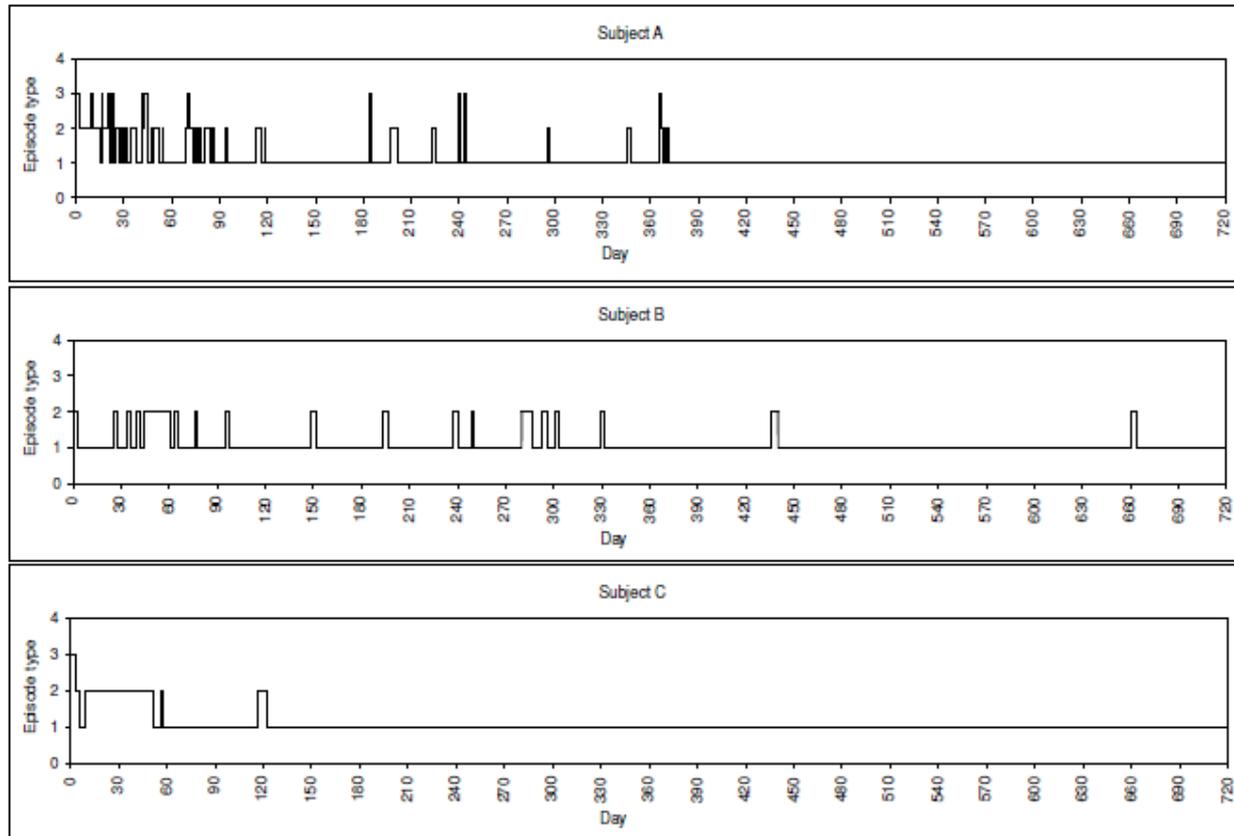


Figure 2: Summaries based on “N” episodes from second data set

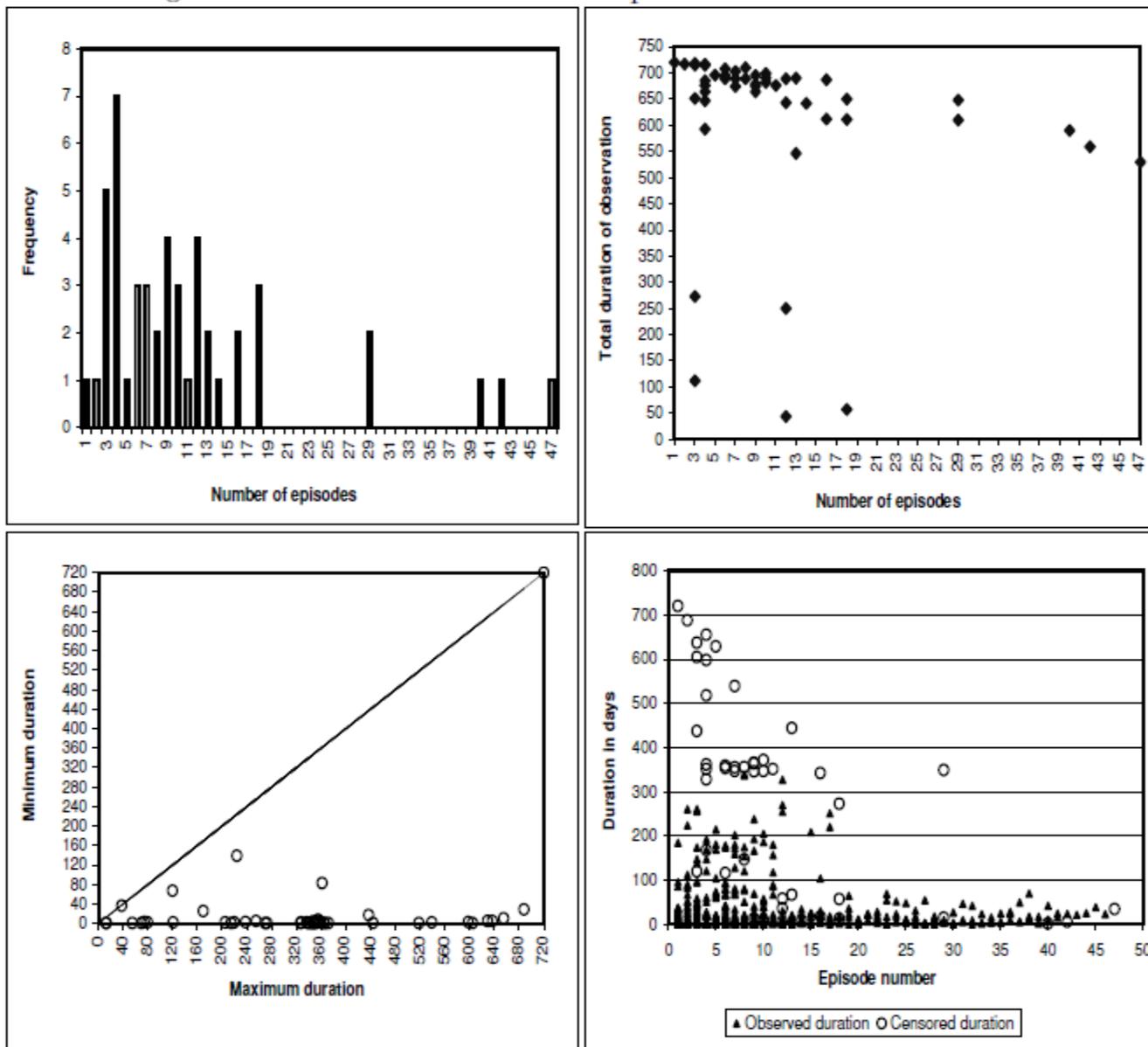
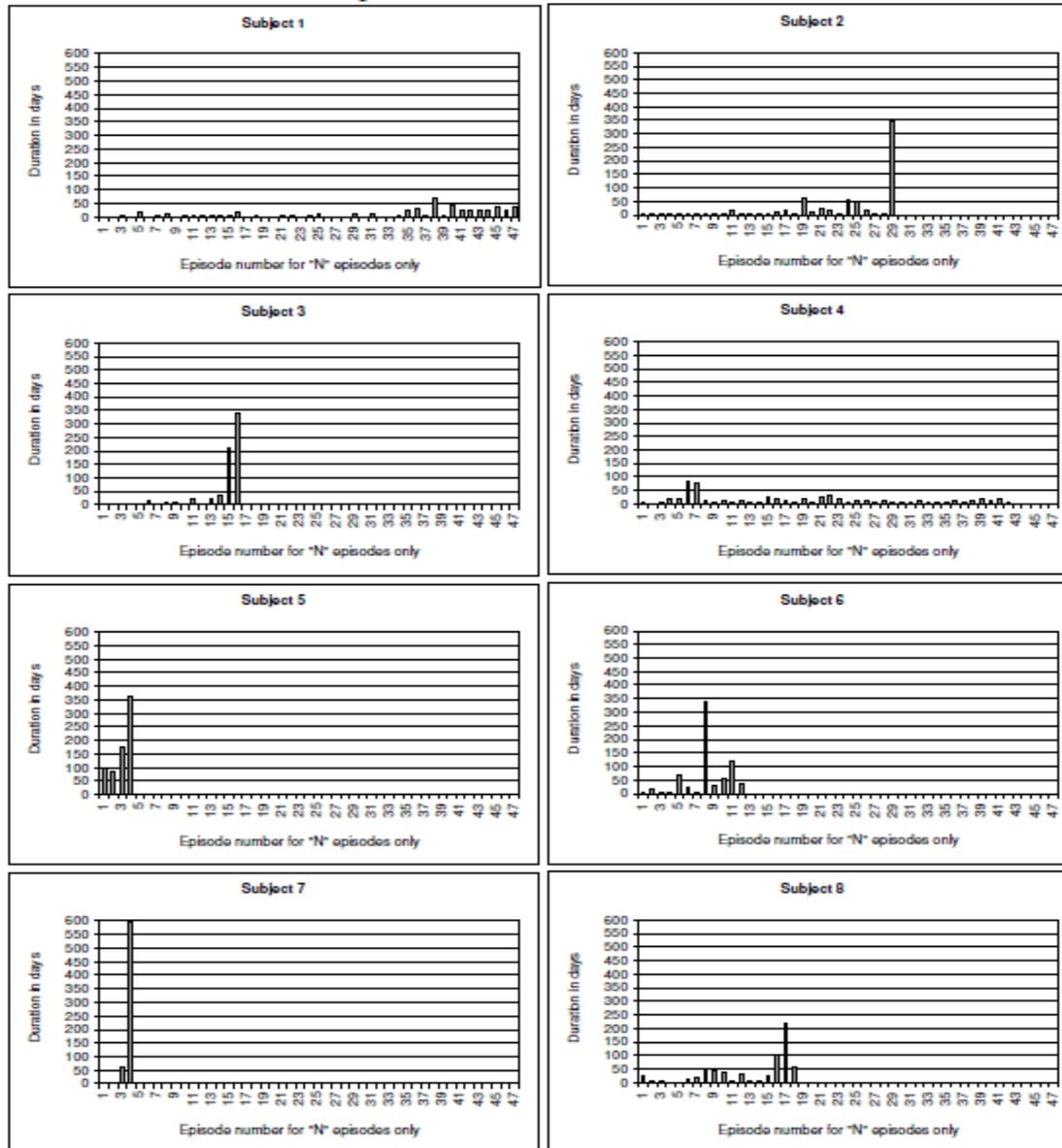


Figure 3: Durations of “N” episodes of some selected individuals from data set-2



# Comments on the data

- High degree of heterogeneity across subjects.
- Large variation in the duration of the “N” episodes within a subject, and not necessarily obvious trend over time.
- However, all treatments are known to have long term benefits and patients are generally expected to experience gradually longer “N” episodes.
- The speed of improvement appears to be different from patient to patient, and it needs to be monitored individually in order to understand the underlying pattern.

# Some simplifying conventions for the modeling

- We lump the possibly several consecutive episodes of bleeding and spotting that may occur between two adjacent non-bleeding episodes into a single combined episode.
- We divide such a combined response further into three types:
  - in a “B” episode there is bleeding, but no spotting, between two adjacent non-bleeding (“N”) periods,
  - in an “S” episode there is only spotting but no bleeding, and
  - in a “BS” episode there are one or more consecutive spells of both bleeding and spotting without an intermittent non-bleeding period.

# Introducing a latent process

- We employ a hierarchical model structure where, in addition to the observed alternating durations in the “N” and “B/S/BS” states, there is a latent process ( $L_t$ ) that would correspond to the woman’s underlying “physiological status” at any given time  $t$ .
- The individually different responses over time to the HRT could then be attributed to individually different sample paths of this underlying process.

# Introducing a latent process (cont'd)

- Considering just four such states for the latent process ( $L_t$ ), and ordering them in an appropriate way, seems to be sufficient for an adequate enough description.
- For this, we denote:
  - “N-4”: a state which has the tendency of leading, on the observed level, to short durations in the “N” phase,
  - “N-3”: a state in which the durations in the “N” phase are (stochastically) somewhat longer,
  - “N-2”: a state in which they are longer still, and
  - “N-1”: a state in which “N” is nearly absorbing.

# Bivariate status process

- Combining the latent process with the observed data, we describe the status of an individual woman at time  $t$  (when measured from the beginning of the treatment) by a pair of random variables, say  $(O_t, L_t)$ .
- Here  $O_t$  is the observed status determined from the available diary data and assuming one of the four possible values “N”, “B”, “S” and “BS”, and  $L_t$  is the corresponding unobserved (latent) status variable with the possible values “N-1”, “N-2”, “N-3” and “N-4” as described above.
- The observed process  $(O_t)$  is alternating in the sense that after every two transitions it has to be back in state “N”.

# Bivariate status process (cont'd)

- We assume that transitions in the latent process ( $L_t$ ) can only happen at times at which the process ( $O_t$ ) returns to the state “N”.
- A “Hidden Markov” structure for the model would not be appropriate: Even with the latent variable  $L_t$  being included as a part of the “current state” description at time  $t$ , its knowledge together with  $O_t$  would not provide an adequate basis for predicting the future behavior of the observed process ( $O_t$ ).
- Therefore we need to extend the memory from the current state of the process ( $O_t, L_t$ ) to include also some relevant aspect of its history.

# Second order memory structure

- Let  $0 = T_0 < T_1 < T_2 < \dots$  be the times at which the observed status  $O_t$  of an individual woman changes, and let  $X_1, X_2, \dots$  be the random durations  $X_j = T_j - T_{j-1}, j \geq 1$ .
- Define “phase indicators”  $I_1, I_2, \dots$  by:  $I_j = 1$  if the  $j$ -th episode was in an “N” state, and  $I_j = 0$  otherwise.
- Finally, define  $C_1, C_2, \dots$  by:  $C_j = O_{T(j-1)}$  if  $I_j = 0$ , and  $C_j = L_{T(j-1)}$  otherwise.
- Thus the the sequence  $(C_j)$  alternates between an observed and a latent state: During an “N” episode it is identical to the corresponding latent status, and during a bleeding and/or spotting it specifies the type (“B”, “S” or “BS”) of this episode.

## Second order memory structure (cont'd)

- If the woman is currently, at time  $t \in [T_{j-1}, T_j)$ , in an “N” state, so that the corresponding phase indicator then has the value  $I_j = 1$ , we assume that the intensity of leaving that state can depend on the corresponding (latent) state  $C_j$ .
- Furthermore, when this transition into the new state  $C_{j+1}$  (being then one of the states “B”, “S” and “BS”) actually happens, we allow the transition probabilities to depend, in addition to the current (latent) status given by  $C_j$ , also on the (observed) status  $C_{j-1}$  during the immediately preceding bleeding and/or spotting episode.

## Second order memory structure (cont'd)

- If she is currently, at time  $t \in [T_{j-1}, T_j)$ , in one of the observed states “B”, “S” and “BS”, so that the corresponding phase indicator then has the value  $I_j = 0$ , we assume that the intensity of leaving that state can depend on the current (observed) state  $C_j$ .
- Furthermore, when this transition into the new state  $C_{j+1}$  (being then one of the latent states “N-1”, ..., “N-4”) actually happens, we allow the transition probabilities to depend, in addition to the current (observed) status given by  $C_j$ , also on the corresponding (latent) status given by  $C_{j-1}$ .

- (1) Duration  $j$  of subject  $i$ :  $(X_{i,j} | Tr_i, I_{i,j}, C_{i,j}, \beta(Tr_i, I_{i,j}, C_{i,j})) \sim Exp(\beta(Tr_i, I_{i,j}, C_{i,j}))$ ,  
where  $i = 1, \dots, 163$  and  $j \geq 1$ ,
- (2) Therapy variables (observed):  $Tr_i = 1, 2, 3$  indicating type of HRT given to subject  $i$
- (3) Phase variables:  $I_{i,1} | Tr_i \sim Bernoulli(\mathbf{P}_{00}(Tr_i))$  and  
 $I_{i,j} = 1 - I_{i,j-1}$  where  $i = 1, \dots, 163$  and  $j \geq 2$ ,
- (4) Partially latent state variables:  $(C_{i,1} | Tr_i, I_{i,1} \mathbf{P}_0) \sim Multinomial(1, \mathbf{P}_0(Tr_i, I_{i,1}))$ ,  
 $(C_{i,2} | Tr_i, I_{i,2} \mathbf{P}_0) \sim Multinomial(1, \mathbf{P}_0(Tr_i, I_{i,2}))$ ,  
 $(C_{i,j} | Tr_i, I_{i,j}, C_{i,j-1}, C_{i,j-2}, \mathbf{P})$   
 $\sim Multinomial(1, \mathbf{P}(Tr_i, I_{i,j}, C_{i,j-1}, C_{i,j-2}))$ ,  
where  $i = 1, \dots, 163$  and  $j \geq 3$
- (5) State specific hazards:  $\beta_{0,l} \sim Gamma(0.1, 0.1)$ , where  $l = 1, 2, 3$  and  
 $\beta_{1,1} \equiv 0.00001$ ,  
 $\beta_{1,l} \sim Gamma(0.1, 0.1)$ , where  $l = 2, \dots, k$  and  
 $\beta_{1,1} \leq \beta_{1,2} \leq \dots \leq \beta_{1,k-1} \leq \beta_{1,k}$ ,
- (6) Transition probabilities:  $\mathbf{P}(l, 0, k_1, k_2) \sim Dirichlet(I_0)$ ,  
where  $l = 1, 2, 3, k_1 = 1, \dots, k$ , and  $k_2 = 1, 2, 3$ ,  
 $\mathbf{P}(l, 1, k_1, k_2) \sim Dirichlet(I_1)$ ,  
where  $l = 1, 2, 3, k_1 = 1, 2, 3$ , and  $k_2 = 1, \dots, k$
- (7) Initial distribution:  $\mathbf{P}_0(l, 0) \sim Dirichlet(1_0)$   
 $\mathbf{P}_0(l, 1) \sim Dirichlet(1_1)$   
where  $1_0 = (1)_{3 \times 1}$  and  $1_1 = (1)_{k \times 1}$ ,
- (8) Phase probabilities:  $\mathbf{P}_{00}(l) \sim Uniform(0, 1)$  where  $l = 1, 2, 3$ .

# Numerical implementation

- The numerical estimation of the desired model parameters and other unknown quantities was carried out using Markov Chain Monte Carlo (MCMC) method, by drawing a large MCMC sample from the posterior distributions of the parameters.
- The hierarchical Bayesian model was specified in BUGS language and implemented in WinBUGS.

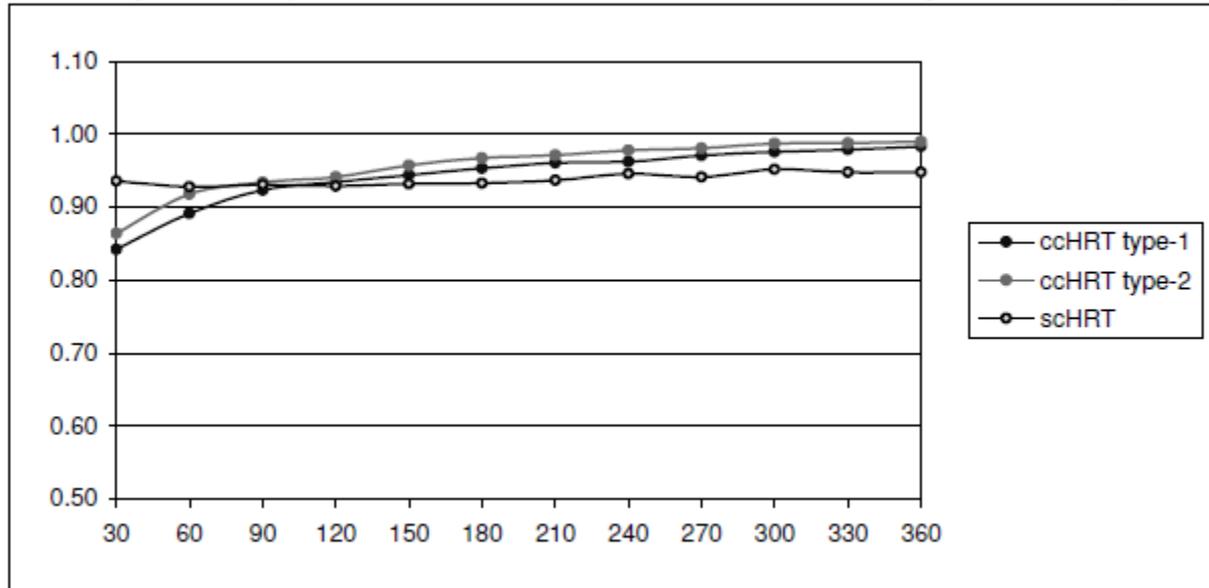
# Some estimates of on model parameters

Table 5: HRT and phase specific estimates posterior means of hazards of leaving the current state:

HRT	Phase type N (latent) same for all three HRTs				Phase type S/SB/B (observed)		
	1	2	3	4	S	SB	B
ccHRT type-1	0.00	0.02	0.04	0.10	0.19	0.13	0.75
ccHRT type-2					0.24	0.16	0.29
scHRT					0.33	0.44	0.24

# Predicted treatment effects

Figure 4: Predicted probability of “N” state or otherwise at fixed time points for a generic individual



- Predicted treatment effects (cont'd)

Table 8: HRT specific predicted time to cumulative amenorrhea for a generic individual:

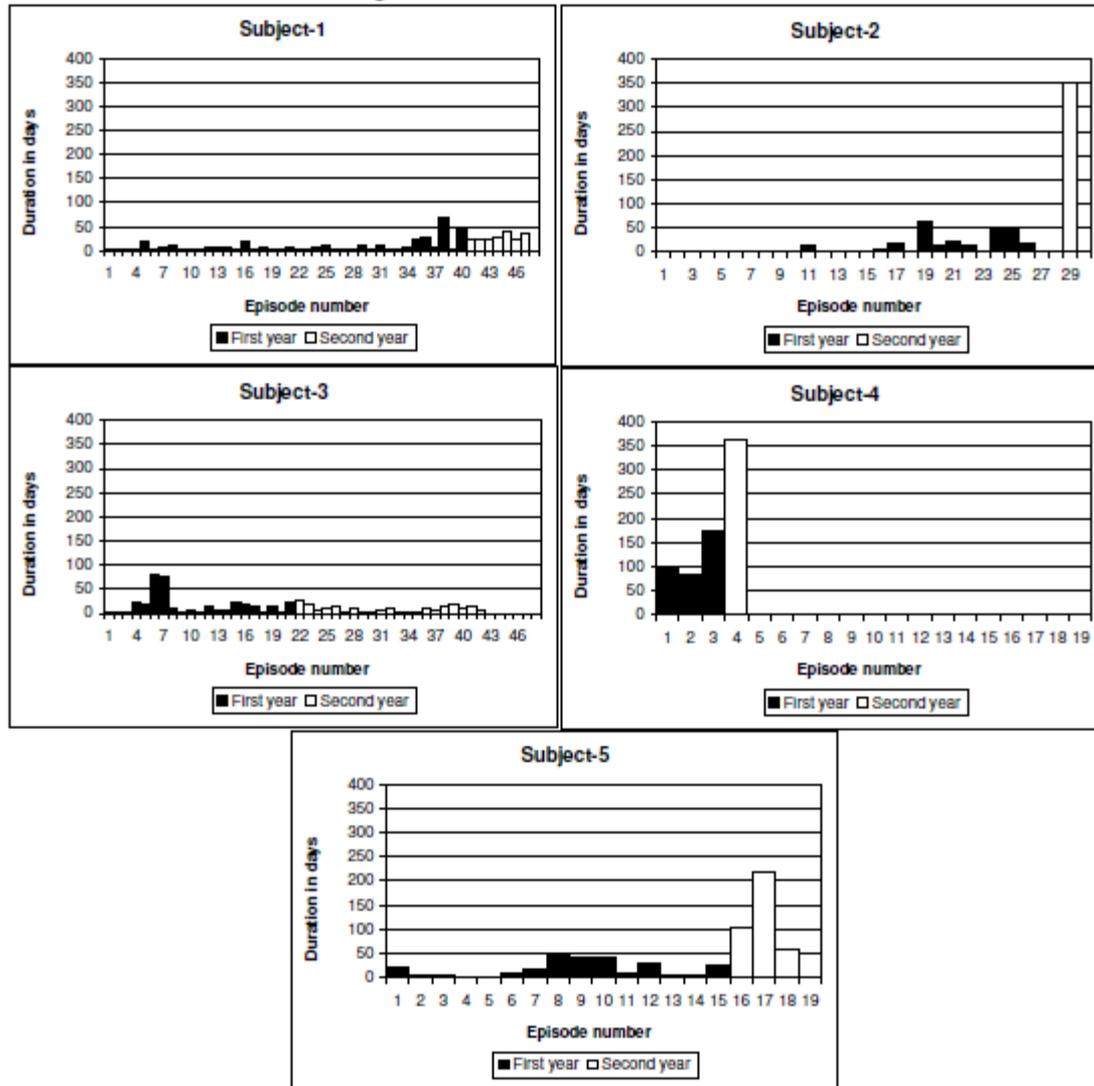
HRT	Predicted median	Predicted mean
ccHRT type-1	99.8	153.1
ccHRT type-2	75.1	121.4
scHRT	327.7	535.7

# Predicted treatment effects (cont'd)

Table 9: HRT specific predicted proportions of time spent in different states till cumulative amenorrhea for a generic individual:

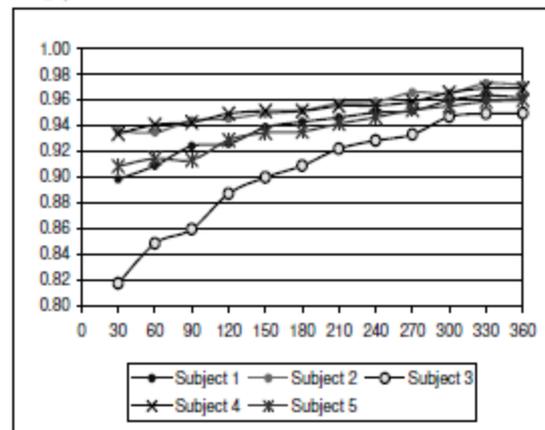
HRT	Phase type N				Phase type S/SB/B		
	1	2	3	4	S	SB	B
ccHRT type-1	0.00	0.48	0.21	0.13	0.11	0.06	0.01
ccHRT type-2	0.00	0.47	0.20	0.15	0.11	0.05	0.02
scHRT	0.00	0.23	0.56	0.07	0.03	0.04	0.01

Figure 5: Durations of “N” episodes for five selected individuals from data set 2



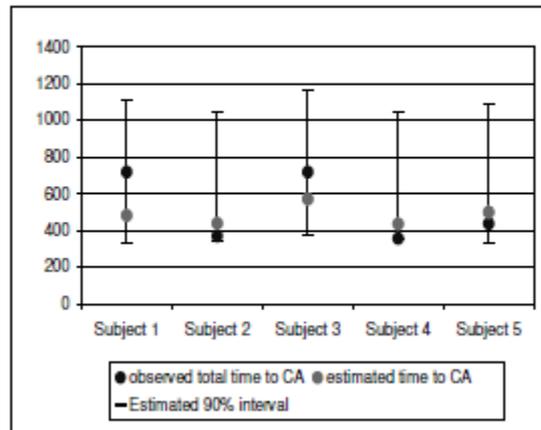
# Predicted treatment effects (cont'd)

Figure 6: Predictive probabilities of being in state “N” at different time points in the second year (based on data from first year only) for 5 individuals from data set 2



# Predicted treatment effects (cont'd)

Figure 7: Posterior predictive 90% interval of time to cumulative amenorrhea (CA) and actually observed/censored at the end of observation interval for 5 individuals from data set



# Reference (unpublished research report):

- Modelling and analysis of time in-homogeneous recurrent event processes in heterogeneous population: A case study of HRTs

Madhuchhanda Bhattacharjee and Elja Arjas

Rolf Nevanlinna Institute

University of Helsinki

Finland

February 8, 2005

# Thanks to

- Madhu
- Leiras/Schering/Bayer Finland