Análisis de datos longitudinales

Grado en Estadística

Tema 2 – Sesión 6

Análisis de Supervivencia

Eventos recurrentes (II)

Juan R González

Departamento de Matemáticas, UAB

Instituto de Salud Global Barcelona (ISGlobal)

Outlined talk

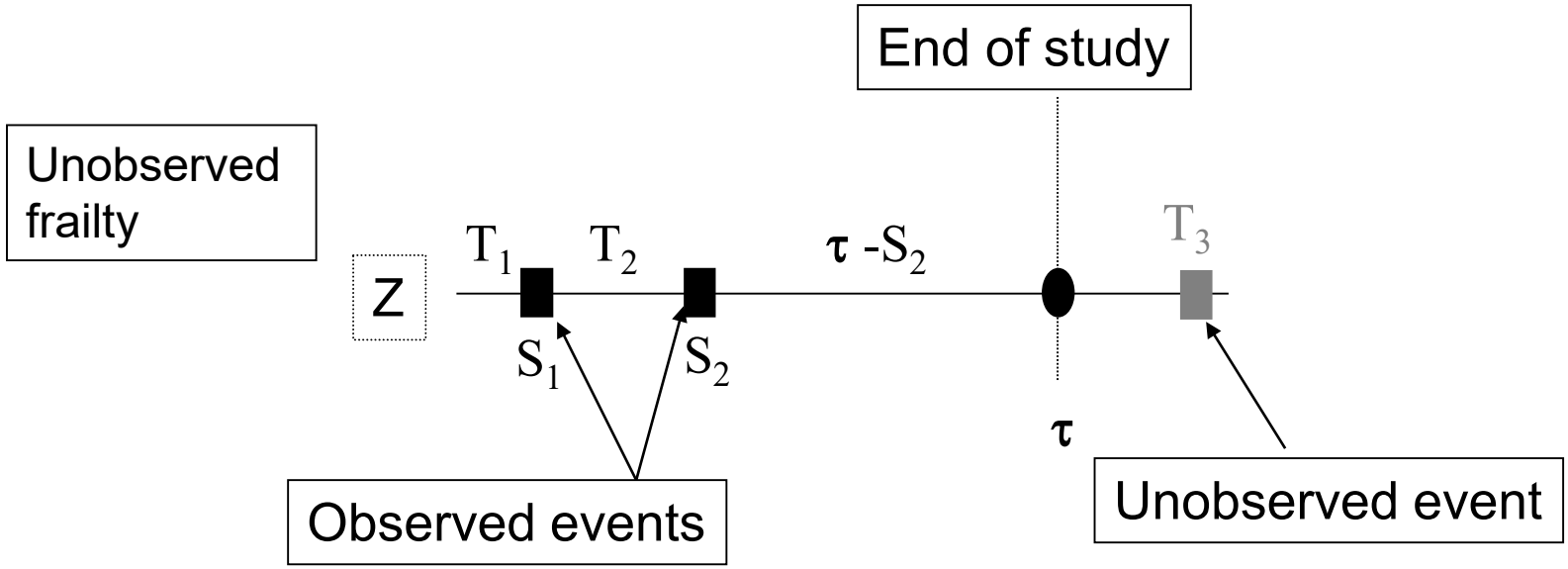
- Introduction: data description
- Repeated events features
  - ‘At risk’ formulation
  - Within-subject correlation
- Existing Cox extension models
- Model selection

Recurrent phenomena

- In Public Health and Medical Settings
  - hospitalization of a subject with a chronic disease
  - tumor occurrence
  - cyclic movements in the small bowel during fasting state
  - episodes of depression

- In Reliability, Engineering, and Economic Settings
  - failure of a mechanical/electronic system
  - warranty claims
  - Dow Jones index changes by more than 200 points
  - occurrence of a certain type of accident (nuclear)

Random entities: one subject

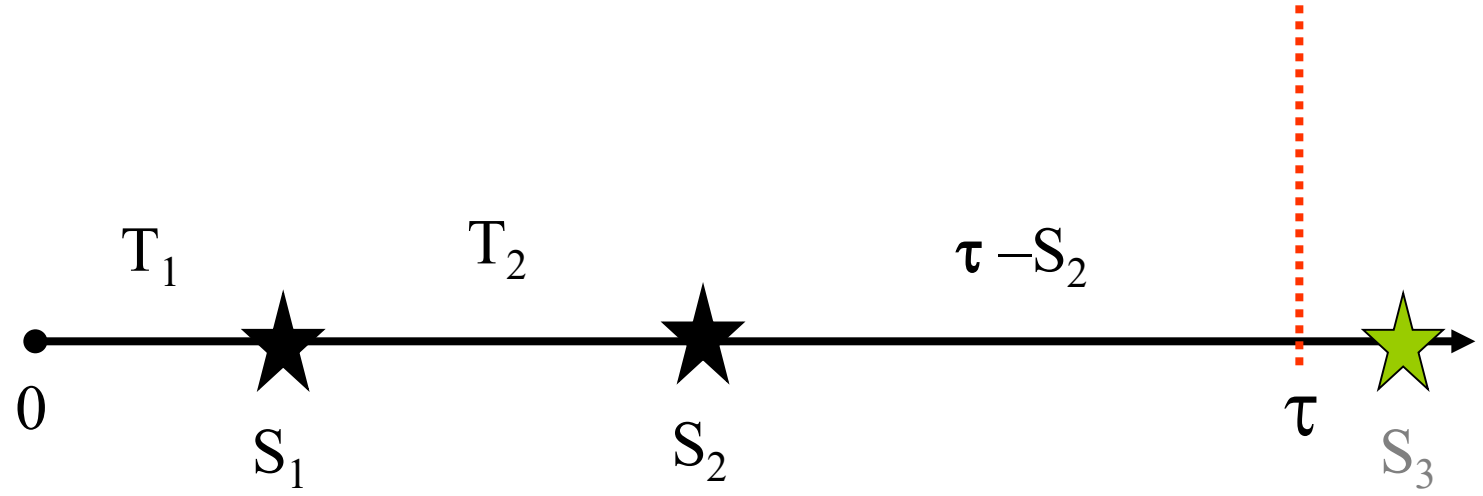


- $T_1, T_2, T_3, \ldots$ = inter-event or gap times
- $S_1, S_2, S_3, \ldots$ = calendar times of event occurrences
- $X(s)$ = covariate vector, possibly time-dependent
- τ = end of observation period
- Accrued History: $F^{\dagger} = \{F^{\dagger}(s) : s \geq 0\}$
- Z = unobserved frailty variable
- $N^{\dagger}(s)$ = number of events in $[0, s]$
- $Y^{\dagger}(s)$ = at-risk indicator at time s

Repeated events

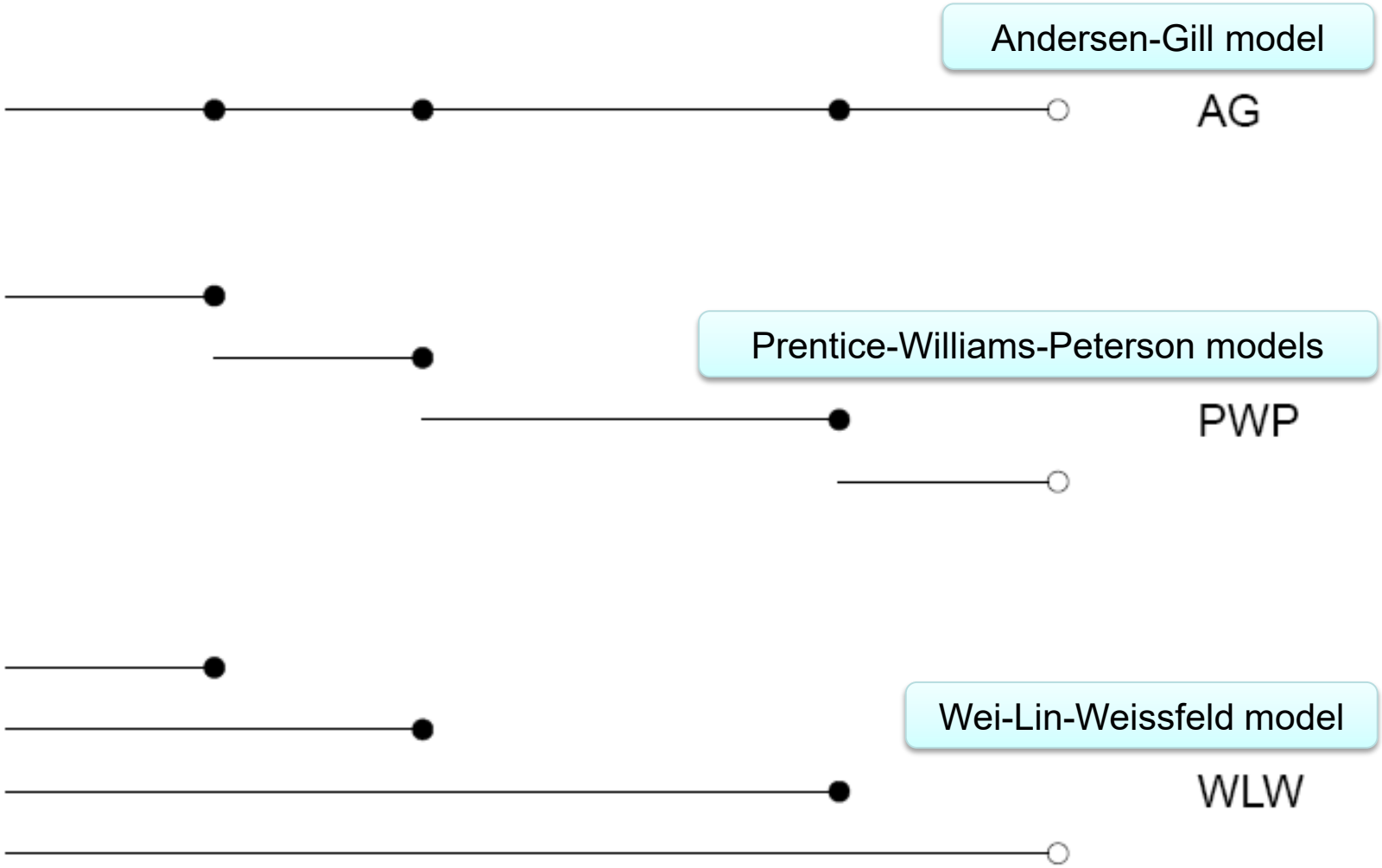
- At risk process
  - Gap time
  - Calendar time
- Within-subject correlation (no i.i.d.)
  - Heterogeneity across individuals
  - Event dependence

‘At-risk processes’

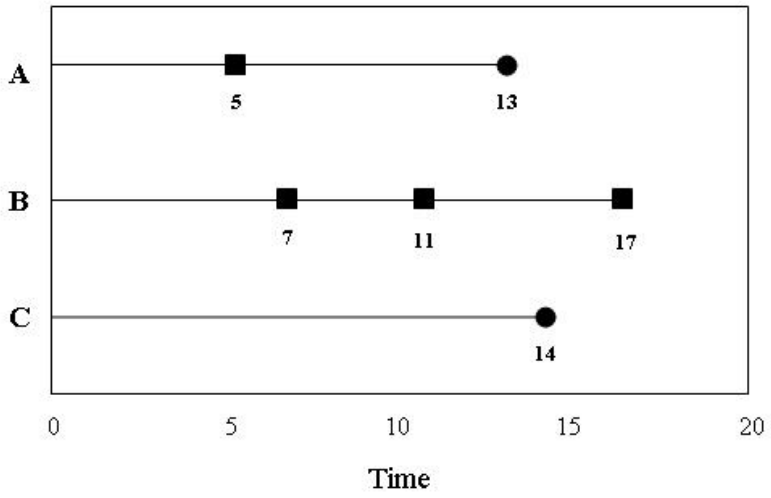


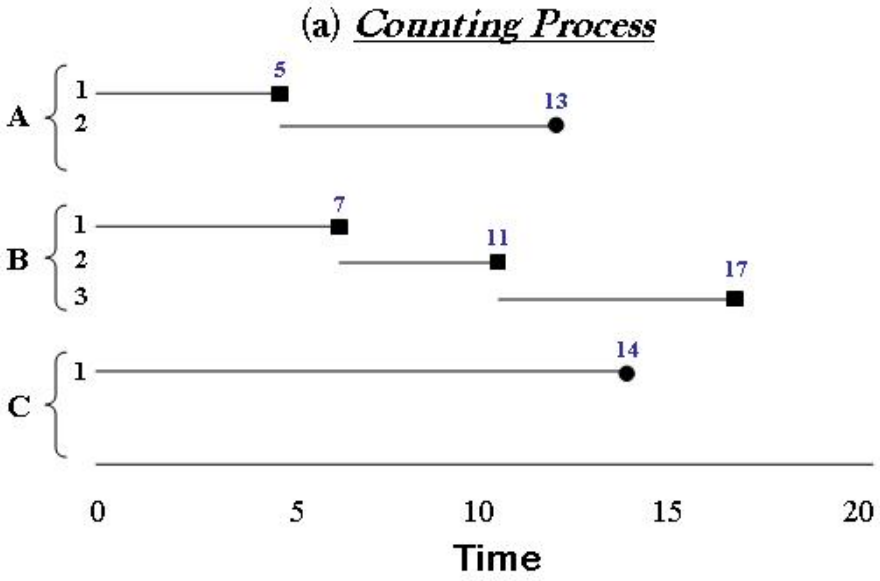
- When a subject is at risk
  - Interoccurrence time (Gap time): $T_i = S_i - S_{i-1}$
  - Calendar time: $S_i = T_1 + T_2 + \ldots + T_i$

‘At-risk processes’

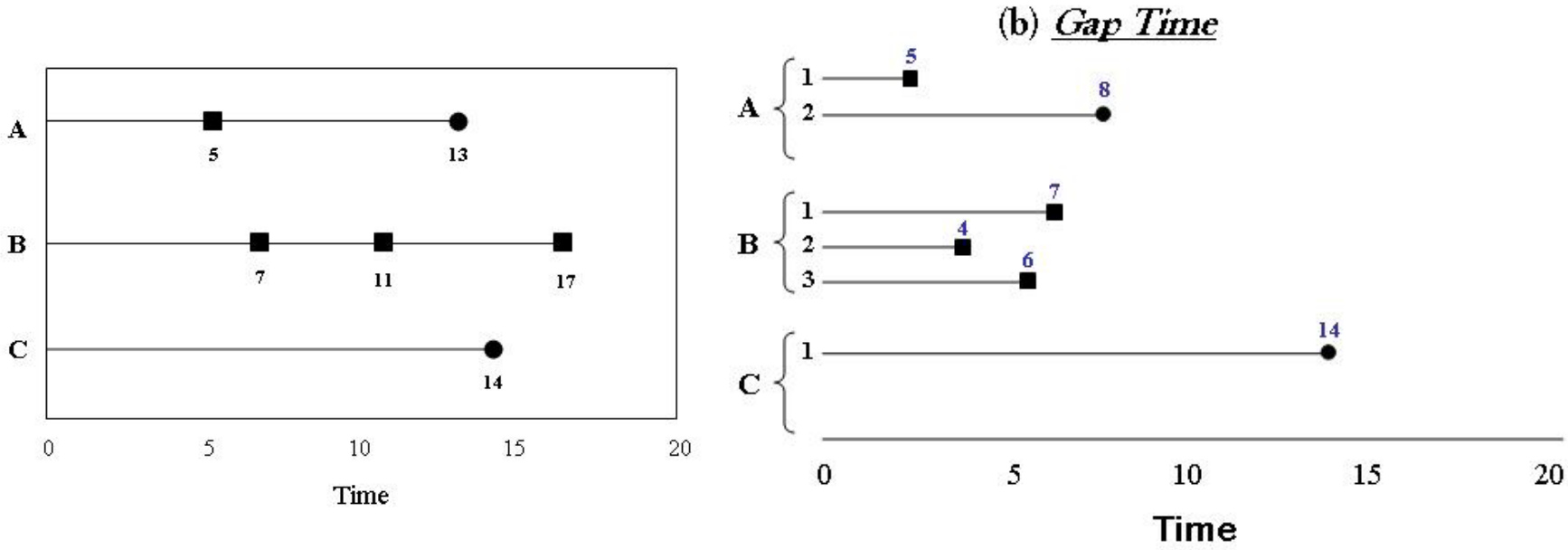


‘At-risk processes’

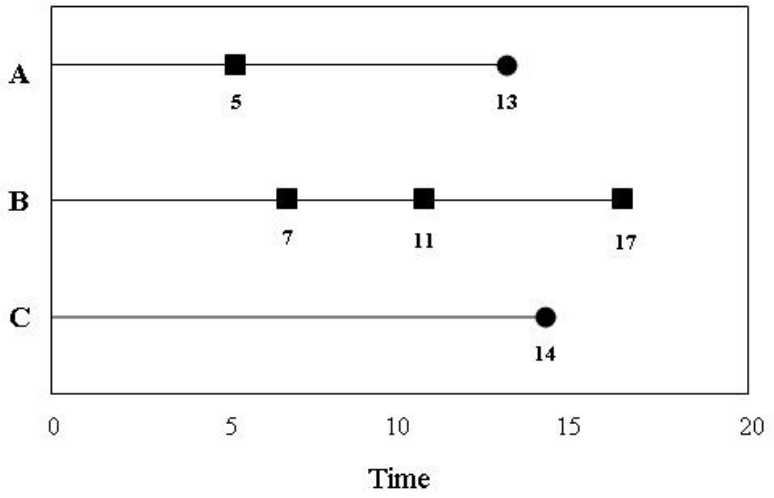


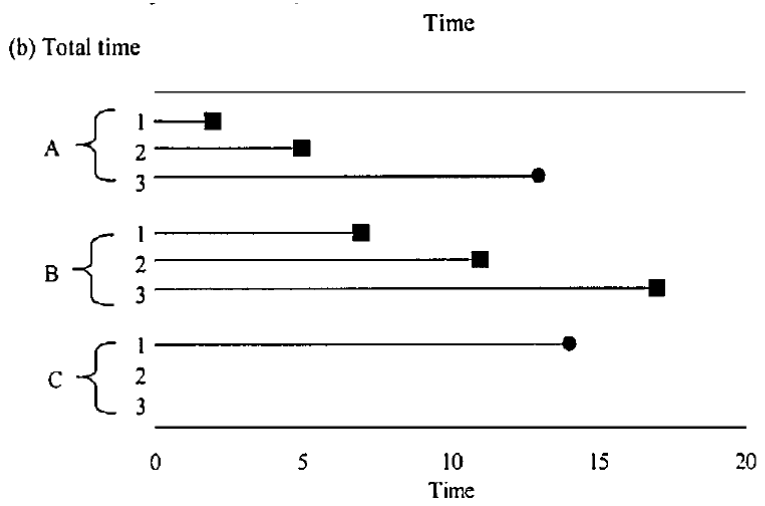


‘At-risk processes’

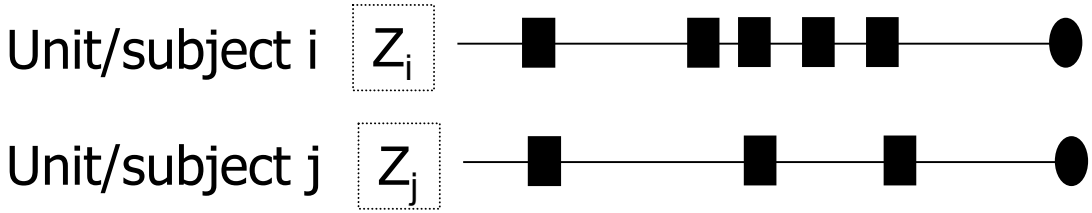


‘At-risk processes’





Within-subject correlation



- Biomedical data (uncontrolled variables, non-measurable variables: genetic susceptibility, …)
- F non i.i.d.
- There exists a random variable Z with known distribution. If we condition to Z=z the interocurrence times are i.i.d.

- Approaches:
  - Variance-corrected models
  - Frailty models (siguiente sesión)

Variance-corrected models

- Use extensions of Cox model
- Idea: Variance independent across groups (or individuals) but not necessarily within groups
- Robust variance estimatior: 'sandwich' estimate

$$V = I^{-1} B I^{-1}$$

B: correction factor, jacknife (Therneau and Hamilton '97)

I: information matrix

Cox model for censored data

- Cox proportional hazard model

$$\lambda_i(t; X_i) = \lambda_0(t) e^{\beta' X_i(t)}$$

$\lambda_0(t)$ Baseline hazard

X_i Vector of covariates

Cox-based models

- Andersen-Gill model:

- The Andersen-Gill model is probably the most often applied model for recurrent event times and is a simple extension of the Cox model.
- It is based on the assumption that the instantaneous risk to experience an event at time *t* since study entry remains the same irrespective of the fact whether previous events occurred or not (strong assumption).
- The Andersen-Gill model therefore aims at estimating the same quantity as the common Cox model given by the all-cause hazard ratio.
- Allow for heterogeneity using variance-corrected estimator:

$$V = I^{-1} B I^{-1}$$

B, e.g. jackkniffe (Therneau '97)

- If the assumption of independent recurrent event times is not fulfilled, the Anderson-Gill model might still be applied but no longer estimates the all-cause hazard ratio. Instead, the resulting treatment effect estimator is given as a hazard ratio combining direct and indirect effects

Cox-based models

- Prentice, Williams and Peterson (PWP) **conditional models** :

$$\lambda_{ik}(t; X_{ik}) = \lambda_{0K}(t) e^{\beta' X_{ik}(t)}$$

- 'at-risk process' for *jth* event only becomes 1 after the *(j - 1)th* event (e.g., conditional) [counting process formulation: start – stop]
- Allow for heterogeneity using variance-corrected estimator:

$$V = I^{-1} B I^{-1}$$

B, e.g. jackkniffe (Therneau '97)

- Allow for event dependence through stratification
- It can be seen as a Cox model but for each recurrent event a separate hazard function is modeled with an own baseline hazard

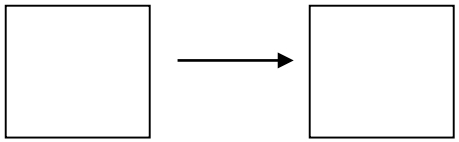
Cox-based models

- Wei, Lin Weissfeld (WLW) **marginal model** :

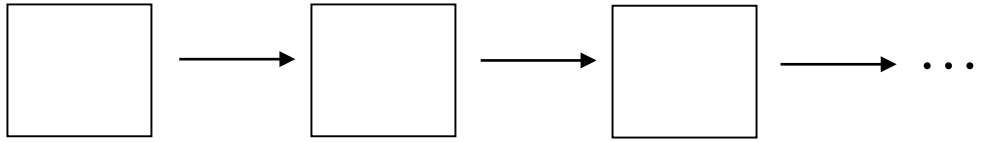
$$\lambda_{ik}(t; X_{ik}) = \lambda_{0k}(t) e^{\beta' X_{ik}(t)}$$

  - Is also an stratified cox model as PWP ones
  - In contrast to the models by PWP, an individual is at risk for every (recurrent) event as long as it is under observation (A subject can be at risk for event *k* before event *k-1* occurs)
  - Allow for heterogeneity using variance-corrected estimator
  - Allow for event dependence through stratification: Different baseline hazard function for each event

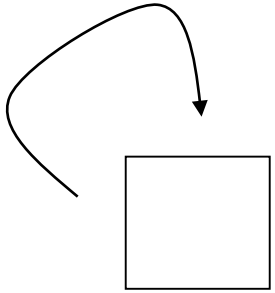
Cox-based models



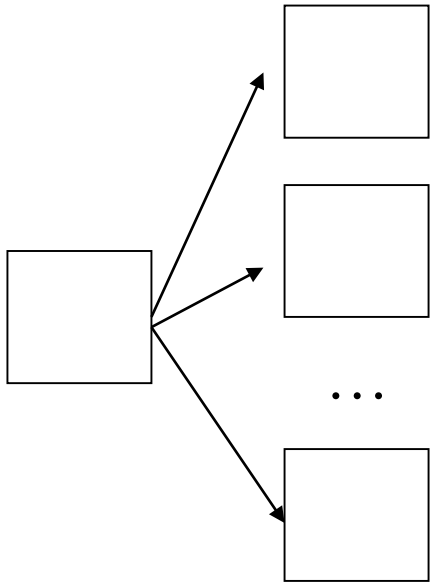
Time to first event



Conditional model

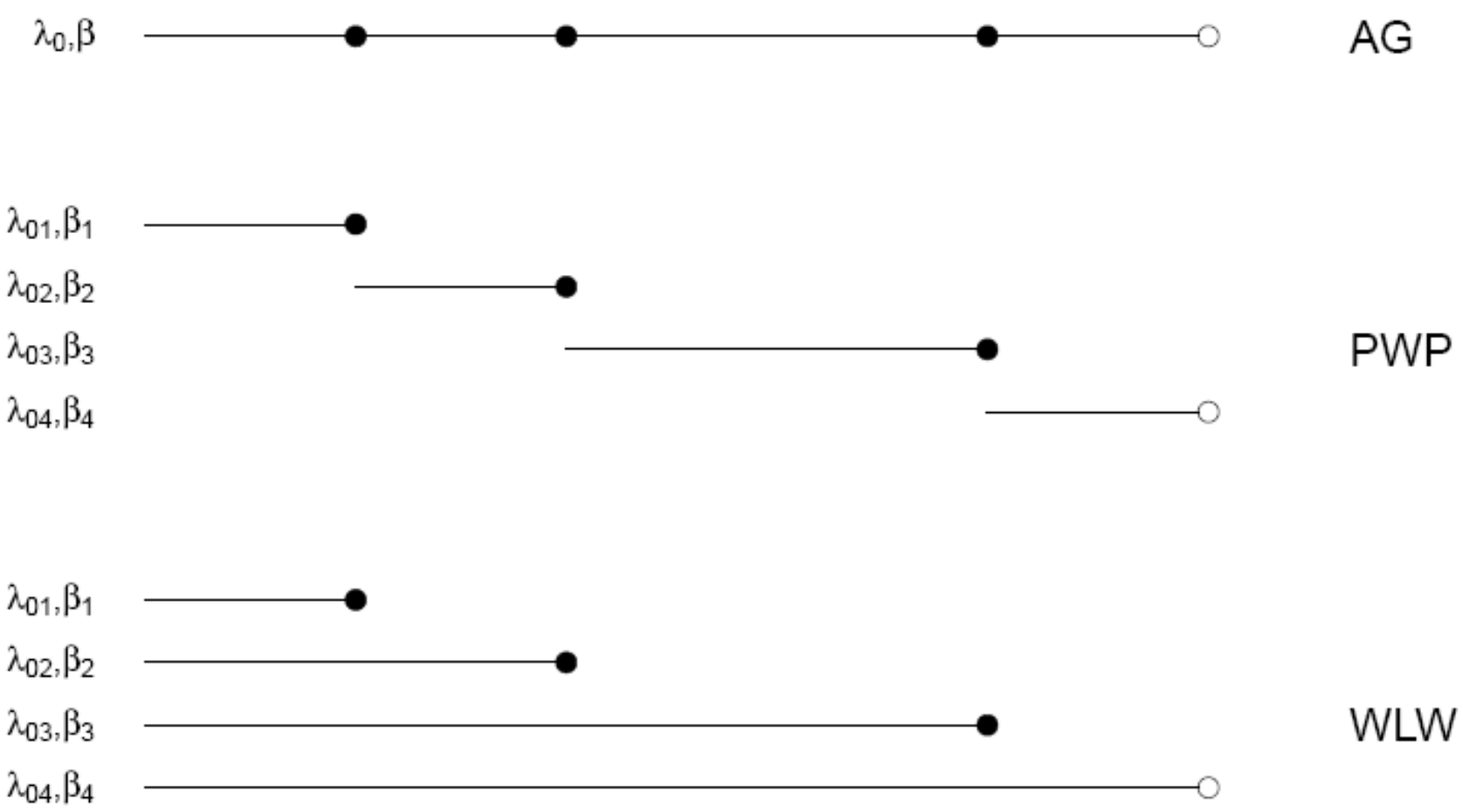


Andersen-Gill model



Marginal model

Cox-based models



Model selection

Consideraciones en el proceso biológico de la enfermedad

- Si suponemos que recidivas consecutivas tienen el mismo riesgo, el modelo AG será adecuado.

- Si consideramos que después de experimentar la primera recidiva, el riesgo de la siguiente se incrementa, sugerirá el uso del modelo PWP.

- Si consideramos que despues de experimentar cada recidiva, el riesgo de la siguiente vuelve a estar como al principo, surgerirá el uso del modelo WLW

Model selection

Consideraciones estadísticas (modelo PWP)

- El uso de la estratificación dependiente en el tiempo, significa que la función de riesgo subyacente puede variar de suceso a suceso, no como en el modelo AG que asume que todos los sucesos son idénticos.

- El modelo PWP define los intervalos de riesgo utilizando el proceso de conteo. Trata los datos como resultados ordenados y cada sujeto se representa como una serie de observaciones (entrada, primer evento], (primer evento, segundo evento],...,(k-ésimo evento, último seguimiento].

Model selection

Consideraciones estadísticas (modelo WLW)

- Prentice, Williams y Peterson proponen un modelo alternativo con intervalo de riesgo el "intervalo de tiempo" para procesos de renovación con intervalos $(0, t_1]$, $(0, t_2 - t_1]$,. . ., donde t_1 denota el tiempo hasta la primera recidiva, t_2 el tiempo hasta la segunda,...etc, se corresponde con la escala de tiempo "tiempo desde la entrada (total time) o desde el último evento (gap time)".