Models for patients' recruitment

Workshop "Modélisation et simulation d'essais cliniques"

Cancéropôle Grand Sud-Ouest

Bordeaux • Limoges • Montpellier • Nîmes • Toulouse

Thuesday 09 April 2015

Nicolas SAVY

INSTITUT de MATHEMATIQUES de TOULOUSE

Mathematics Institute of Toulouse

Institut de Recherche en Santé Publique GIS-IReSP

Research granted by IRESP

- Clinical trials is one of the main elements of the marketing authorization of a new drug
- Such a request has to follow a protocol specifying
  - Patients inclusion and exclusion criteria
  - Statistic analysis plan especially :
    - which test is used
    - what are the type I and type II risks
    - **necessary sample size** N
- In order to recruit these N patients, several investigators centres are involved

Definition

The **recruitment period** is the duration between the initiation of the first of the C investigator centres and the instant $T(N)$ when the N patients are included.

- N is fixed but $T(N)$ is a random variable

What is recruitment period ?

- Clinical trials is one of the main elements of the marketing authorization of a new drug
- Such a request has to follow a protocol specifying
  - Patients inclusion and exclusion criteria
  - Statistic analysis plan especially :
    - which test is used
    - what are the type I and type II risks
    - **necessary sample size** N
- In order to recruit these N patients, several investigators centres are involved

Definition

The **recruitment period** is the duration between the initiation of the first of the C investigator centres and the instant $T(N)$ when the N patients are included.

- N is fixed but $T(N)$ is a random variable

What is recruitment period ?

- Clinical trials is one of the main elements of the marketing authorization of a new drug
- Such a request has to follow a protocol specifying
  - Patients inclusion and exclusion criteria
  - Statistic analysis plan especially :
    - which test is used
    - what are the type I and type II risks
    - **necessary sample size N**
- In order to recruit these N patients, several investigators centres are involved

Definition

The **recruitment period** is the duration between the initiation of the first of the C investigator centres and the instant $T(N)$ when the N patients are included.

- N is fixed but $T(N)$ is a random variable

- Clinical trials is one of the main elements of the marketing authorization of a new drug
- Such a request has to follow a protocol specifying
  - Patients inclusion and exclusion criteria
  - Statistic analysis plan especially :
    - which test is used
    - what are the type I and type II risks
    - **necessary sample size N**
- In order to recruit these N patients, several investigators centres are involved

Definition

The **recruitment period** is the duration between the initiation of the first of the C investigator centres and the instant $T(N)$ when the N patients are included.

- N is fixed but $T(N)$ is a random variable

What is recruitment period ?

- Clinical trials is one of the main elements of the marketing authorization of a new drug
- Such a request has to follow a protocol specifying
  - Patients inclusion and exclusion criteria
  - Statistic analysis plan especially :
    - which test is used
    - what are the type I and type II risks
    - **necessary sample size** N
- In order to recruit these N patients, several investigators centres are involved

Definition

The **recruitment period** is the duration between the initiation of the first of the C investigator centres and the instant $T(N)$ when the N patients are included.

- N is fixed but $T(N)$ is a random variable

Motivation of those investigations

- **Why a model of recruitment period ?**
  - The duration of the recruitment period is very **hard to control**
  - A clinical trial is **expensive**
    - $ 150.000.000 : Average out-of-pocket clinical cost for each new drug
  - Pharma-Companies need tools to **be able** to decide :
    - to overpass the targeted duration of the trial T_R
    - stop the trial if it is too long

- **What a model of recruitment for ?**
  - To develop tools for the study the feasibility of a clinical trial
    - based on the estimation of $T(N)$ (punctually and by means of CI)
  - To Detect critical point in the recruitment
  - To define decision rules on the recruitment process to reach T_R
    - based on the estimation of the recruitment rate
    - based on the estimation of the number of centre to open

Motivation of those investigations

- **Why a model of recruitment period ?**
  - The duration of the recruitment period is very **hard to control**
  - A clinical trial is **expensive**
    - $ 150.000.000 : Average out-of-pocket clinical cost for each new drug
  - Pharma-Companies need tools to **be able** to decide :
    - to overpass the targeted duration of the trial T_R
    - stop the trial if it is too long

- **What a model of recruitment for ?**
  - To develop tools for the study the feasibility of a clinical trial
    - based on the estimation of $T(N)$ (punctually and by means of CI)
  - To Detect critical point in the recruitment
  - To define decision rules on the recruitment process to reach T_R
    - based on the estimation of the recruitment rate
    - based on the estimation of the number of centre to open

- **How to model the recruitment period ?**
  - Analogy with queueing theory

Queueing theory		**Clinical research**
Storage capacity	$\longleftrightarrow$	target population or cohort
Server	$\longleftrightarrow$	None
Exit process	$\longleftrightarrow$	Drop-out patients
Entry process	$\longleftrightarrow$	Recruitment

- It is thus natural to model the recruitment period by means of **Poisson processes**.

- **How to model the recruitment period ?**
  - Analogy with queueing theory

Queueing theory		**Clinical research**
Storage capacity	$\longleftrightarrow$	target population or cohort
Server	$\longleftrightarrow$	None
Exit process	$\longleftrightarrow$	Drop-out patients
Entry process	$\longleftrightarrow$	Recruitment

- It is thus natural to model the recruitment period by means of **Poisson processes**.

The recruitment process model

Consider a multicentric trial involving C investigator centres

- N : number of patients to be recruited
- T_R : expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf \{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

Consider a multicentric trial involving C investigator centres

- N : number of patients to be recruited
- T_R : expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

The recruitment process model

Consider a multicentric trial involving C investigator centres

- N : number of patients to be recruited
- T_R : expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

The recruitment process model

Consider a multicentric trial involving C investigator centres

- N : number of patients to be recruited
- T_R : expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

The recruitment process model

INSTITUT de MATHEMATIQUES de TOULOUSE

Consider a multicentric trial involving C investigator centres

- N : number of patients to be recruited
- T_R : expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf \{t \in \mathbb{R} \mid \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

The recruitment process model

Consider a multicentric trial involving C investigator centres

- N : number of patients to be recruited
- T_R : expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

The recruitment process model

Consider a multicentric trial involving C investigator centres

- N : number of patients to be recruited
- T_R : expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

Consider a multicentric trial involving C investigator centres

- N : number of patients to be recruited
- T_R : expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf \{t \in \mathbb{R} \mid \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

The recruitment process model

Consider a multicentric trial involving C investigator centres

- N : number of patients to be recruited
- T_R : expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf \{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

Consider a multicentric trial involving C investigator centres

- N : number of patients to be recruited
- T_R : expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

Theorem

- ***If λ is known (given by the investigator)*** *then*

  The ***feasibility of the trial*** *expresses by :*

$$\mathbb{P}\left[\mathcal{N}(T_R) \geq N \mid \mathcal{F}_{T_1}\right] = 1 - \sum_{k=0}^{N-N_1-1} \frac{1}{k!} \int_{\mathbb{R}^C} \left(\int_{T_1}^{T_R} (x_1 + \ldots + x_C)dt\right)^k e^{-\int_{T_1}^{T}(x_1+\ldots+x_C)dt} \prod_{i=1}^{C} p_\lambda^{T_1}(x_i)\, dx_i \quad (1)$$

  The ***expected duration*** $\mathbb{E}[T_n]$ *of the trial expresses by :*

$$\mathbb{E}\left[\inf_{t\in\mathbb{R}} \{\mathcal{N}(t) \geq N\} \mid \mathcal{F}_{T_1}\right] = N \int_{\mathbb{R}^C} \frac{p_\lambda^{T_1}(x_1, \ldots, x_C)}{x_1 + \ldots + x_C} dx_1 \ldots dx_C \quad (2)$$

  Involving $p_\lambda^{T_1}$ the forward density of λ.

- ***If λ is unknown*** *then*
  - *$\hat{\lambda}$ an estimation of λ from the data collected on $[0, T_1]$*
  - *Replace λ by $\hat{\lambda}$ in (1) and (2)*

Theorem

- ***If λ is known (given by the investigator)*** *then*

  *The **feasibility of the trial** expresses by :*

$$\mathbb{P}\left[\mathcal{N}(T_R) \geq N \mid \mathcal{F}_{T_1}\right]$$
$$= 1 - \sum_{k=0}^{N-N_1-1} \frac{1}{k!} \int_{\mathbb{R}^C} \left(\int_{T_1}^{T_R} (x_1 + \ldots + x_C) dt\right)^k e^{-\int_{T_1}^{T} (x_1 + \ldots + x_C) dt} \prod_{i=1}^{C} p_\lambda^{T_1}(x_i)\, dx_i \qquad (1)$$

  *The **expected duration** $\mathbb{E}[T_n]$ of the trial expresses by :*

$$\mathbb{E}\left[\inf_{t \in \mathbb{R}} \{\mathcal{N}(t) \geq N\} \mid \mathcal{F}_{T_1}\right] = N \int_{\mathbb{R}^C} \frac{p_\lambda^{T_1}(x_1, \ldots, x_C)}{x_1 + \ldots + x_C} dx_1 \ldots dx_C \qquad (2)$$

  Involving $p_\lambda^{T_1}$ the forward density of λ.

- ***If λ is unknown*** *then*
  - *$\hat{\lambda}$ an estimation of λ from the data collected on $[0, T_1]$*
  - *Replace λ by $\hat{\lambda}$ in (1) and (2)*

Theorem

- ***If λ is known (given by the investigator)*** *then*

  The ***feasibility of the trial*** *expresses by :*

$$\mathbb{P}\left[\mathcal{N}(T_R) \geq N \mid \mathcal{F}_{T_1}\right] = 1 - \sum_{k=0}^{N-N_1-1} \frac{1}{k!} \int_{\mathbb{R}^C} \left(\int_{T_1}^{T_R} (x_1 + \ldots + x_C)dt\right)^k e^{-\int_{T_1}^{T}(x_1+\ldots+x_C)dt} \prod_{i=1}^{C} p_\lambda^{T_1}(x_i)\, dx_i \quad (1)$$

  The ***expected duration*** $\mathbb{E}[T_n]$ *of the trial expresses by :*

$$\mathbb{E}\left[\inf_{t\in\mathbb{R}} \{\mathcal{N}(t) \geq N\} \mid \mathcal{F}_{T_1}\right] = N \int_{\mathbb{R}^C} \frac{p_\lambda^{T_1}(x_1, \ldots, x_C)}{x_1 + \ldots + x_C} dx_1 \ldots dx_C \quad (2)$$

  Involving $p_\lambda^{T_1}$ the forward density of λ.

- ***If λ is unknown*** *then*
  - *$\hat{\lambda}$ an estimation of λ from the data collected on $[0, T_1]$*
  - *Replace λ by $\hat{\lambda}$ in (1) and (2)*

Theorem

- ***If λ is known (given by the investigator)*** *then*

  *The **feasibility of the trial** expresses by :*

$$\mathbb{P}\left[\mathcal{N}(T_R) \geq N \mid \mathcal{F}_{T_1}\right] = 1 - \sum_{k=0}^{N-N_1-1} \frac{1}{k!} \int_{\mathbb{R}^C} \left(\int_{T_1}^{T_R} (x_1 + \ldots + x_C) dt\right)^k e^{-\int_{T_1}^{T} (x_1+\ldots+x_C)dt} \prod_{i=1}^{C} p_\lambda^{T_1}(x_i)\, dx_i \qquad (1)$$

  *The **expected duration** $\mathbb{E}[T_n]$ of the trial expresses by :*

$$\mathbb{E}\left[\inf_{t\in\mathbb{R}} \{\mathcal{N}(t) \geq N\} \mid \mathcal{F}_{T_1}\right] = N \int_{\mathbb{R}^C} \frac{p_\lambda^{T_1}(x_1, \ldots, x_C)}{x_1 + \ldots + x_C} dx_1 \ldots dx_C \qquad (2)$$

  Involving $p_\lambda^{T_1}$ the forward density of λ.

- ***If λ is unknown*** *then*
  - *$\hat{\lambda}$ an estimation of λ from the data collected on $[0, T_1]$*
  - *Replace λ by $\hat{\lambda}$ in (1) and (2)*

Theorem

- ***If λ is known (given by the investigator)*** *then*

  The ***feasibility of the trial*** *expresses by :*

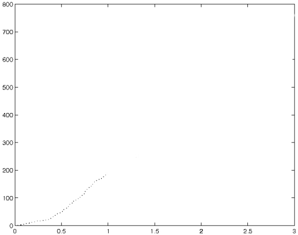
$$\mathbb{P}\left[\mathcal{N}(T_R) \geq N \mid \mathcal{F}_{T_1}\right] = 1 - \sum_{k=0}^{N-N_1-1} \frac{1}{k!} \int_{\mathbb{R}^C} \left(\int_{T_1}^{T_R} (x_1 + \ldots + x_C) dt \right)^k e^{-\int_{T_1}^{T} (x_1 + \ldots + x_C) dt} \prod_{i=1}^{C} p_{\lambda}^{T_1}(x_i)\, dx_i \quad (1)$$

  The ***expected duration*** $\mathbb{E}[T_n]$ *of the trial expresses by :*

$$\mathbb{E}\left[\inf_{t \in \mathbb{R}} \{\mathcal{N}(t) \geq N\} \mid \mathcal{F}_{T_1}\right] = N \int_{\mathbb{R}^C} \frac{p_{\lambda}^{T_1}(x_1, \ldots, x_C)}{x_1 + \ldots + x_C} dx_1 \ldots dx_C \quad (2)$$

  Involving $p_{\lambda}^{T_1}$ *the forward density of* λ.

- ***If λ is unknown*** *then*
  - $\hat{\lambda}$ *an estimation of* λ *from the data collected on* $[0, T_1]$
  - *Replace* λ *by* $\hat{\lambda}$ *in* (1) *and* (2)



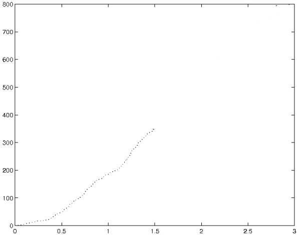


FIGURE: On going study at 1 year (on the left) and at 1.5 year (on the right)

- Dots : Real data used to calibrate the model

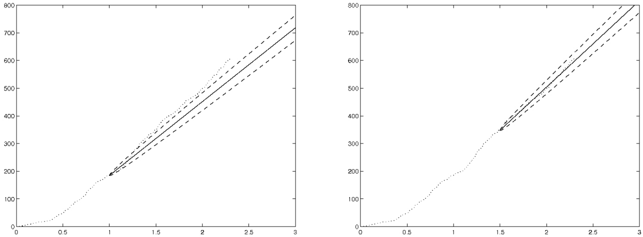


FIGURE: On going study at 1 year (on the left) and at 1.5 year (on the right)

- Dots : Real data used to calibrate the model
- Solid line : estimated number of recruited patients
- Dotted line : Confidence Intervals

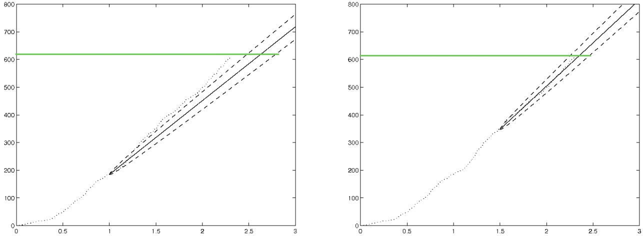
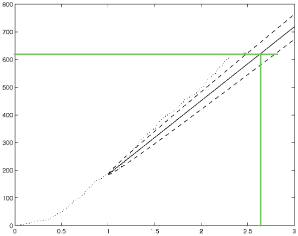


FIGURE: On going study at 1 year (on the left) and at 1.5 year (on the right)

- Dots : Real data used to calibrate the model
- Solid line : estimated number of recruited patients
- Dotted line : Confidence Intervals



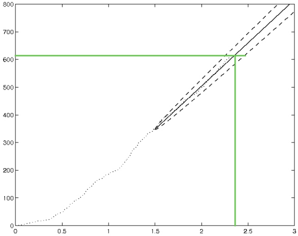


FIGURE: On going study at 1 year (on the left) and at 1.5 year (on the right)

- Dots : Real data used to calibrate the model
- Solid line : estimated number of recruited patients
- Dotted line : Confidence Intervals

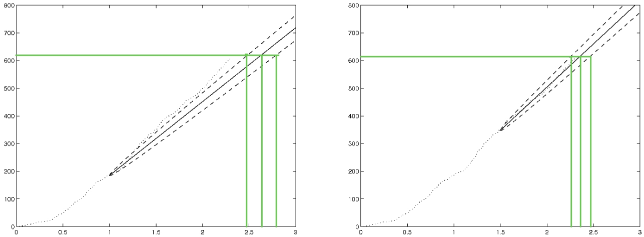


FIGURE: On going study at 1 year (on the left) and at 1.5 year (on the right)

- Dots : Real data used to calibrate the model
- Solid line : estimated number of recruited patients
- Dotted line : Confidence Intervals

Limit of this approach

Problem 1 : If p estimations are needed to describe $\mathcal{N}_i$, $C \cdot p$ estimation are needed to describe $\mathcal{N}$

When C large, this is not relevant

Problem 2 : If centre i has not recruited before T_1, then $\hat{\lambda}_i = 0$ and the model does not authorize centre i to recruit later

Empirical Bayesian model

Ones considers

$$(\lambda_1, \ldots, \lambda_C)$$

is a sample of size C distributed by a certain distribution $\mathcal{L}(\theta)$

Instead of estimate C values of λ, one estimates θ

Limit of this approach

Problem 1 : If p estimations are needed to describe $\mathcal{N}_i$, $C \cdot p$ estimation are needed to describe $\mathcal{N}$

When C large, this is not relevant

Problem 2 : If centre i has not recruited before T_1, then $\hat{\lambda}_i = 0$ and the model does not authorize centre i to recruit later

Empirical Bayesian model

Ones considers

$$(\lambda_1, \ldots, \lambda_C)$$

is a sample of size C distributed by a certain distribution $\mathcal{L}(\theta)$

Instead of estimate C values of λ, one estimates θ

Limit of this approach

Problem 1 : If p estimations are needed to describe $\mathcal{N}_i$, $C \cdot p$ estimation are needed to describe $\mathcal{N}$

When C large, this is not relevant

Problem 2 : If centre i has not recruited before T_1, then $\hat{\lambda}_i = 0$ and the model does not authorize centre i to recruit later

Empirical Bayesian model

Ones considers

$$(\lambda_1, \ldots, \lambda_C)$$

is a sample of size C distributed by a certain distribution $\mathcal{L}(\theta)$

Instead of estimate C values of λ, one estimates θ

Limit of this approach

Problem 1 : If p estimations are needed to describe $\mathcal{N}_i$, $C \cdot p$ estimation are needed to describe $\mathcal{N}$

When C large, this is not relevant

Problem 2 : If centre i has not recruited before T_1, then $\hat{\lambda}_i = 0$ and the model does not authorize centre i to recruit later

Empirical Bayesian model

Ones considers

$$(\lambda_1, \ldots, \lambda_C)$$

is a sample of size C distributed by a certain distribution $\mathcal{L}(\theta)$

Instead of estimate C values of λ, one estimates θ

- **Γ-Poisson model** (Anisimov, Fedorov (2007))
  - Rates are $\Gamma(\alpha, \beta)$ distributed.
  - Distribution of T is explicit.

- **Π-Poisson model** (Mijoule, Savy and **Savy** (2012))
  - Rates are Pareto-(x_m, k_p) distributed.
  - 20% of centres recruit 80% of patients.
  - Distribution of T is no more explicit (Monte Carlo Simulation).

- **$\mathcal{U}\Gamma$-Poisson model** (Mijoule, Savy and **Savy** (2012))
  - Centre opening date are unknown and uniformly distributed

- **Γ-Poisson model** (Anisimov, Fedorov (2007))
  - Rates are $\Gamma(\alpha, \beta)$ distributed.
  - Distribution of T is explicit.
- **Π-Poisson model** (Mijoule, Savy and **Savy** (2012))
  - Rates are Pareto-(x_m, k_p) distributed.
  - 20% of centres recruit 80% of patients.
  - Distribution of T is no more explicit (Monte Carlo Simulation).
- **$\mathcal{U}\Gamma$-Poisson model** (Mijoule, Savy and **Savy** (2012))
  - Centre opening date are unknown and uniformly distributed

- **Γ-Poisson model** (Anisimov, Fedorov (2007))
  - Rates are $\Gamma(\alpha, \beta)$ distributed.
  - Distribution of T is explicit.
- **Π-Poisson model** (Mijoule, Savy and **Savy** (2012))
  - Rates are Pareto-(x_m, k_p) distributed.
  - 20% of centres recruit 80% of patients.
  - Distribution of T is no more explicit (Monte Carlo Simulation).
- **$\mathcal{U}\Gamma$-Poisson model** (Mijoule, Savy and **Savy** (2012))
  - Centre opening date are unknown and uniformly distributed

- **Objectives :**
  - $N = 610$ patients
  - $T_R = 3$ years
  - $C_R = 77$ investigators centres
- **On-going studies :** after 1 year, after 1.5 year and after 2 years
- **The estimated duration** of the trial

The model	Time 1	Time 1.5	Time 2
Constant intensity	3.30	2.63	2.44
Γ-Poisson model	3.31	2.63	2.44
Π-Poisson model	2.63	2.39	2.36
$\mathcal{U}\Gamma$-Poisson model	2.60	2.34	2.36

- **Effective duration** of the trial : **2.31 years**
  - The end of the trial was predicted with an error of **15 days, 10 mouths** before the expected date
  - **56 centres** would be enough for ending in 3 years.

- **Objectives :**
  - $N = 610$ patients
  - $T_R = 3$ years
  - $C_R = 77$ investigators centres
- **On-going studies :** after 1 year, after 1.5 year and after 2 years
- **The estimated duration** of the trial

The model	Time 1	Time 1.5	Time 2
Constant intensity	3.30	2.63	2.44
Γ-Poisson model	3.31	2.63	2.44
Π-Poisson model	2.63	2.39	2.36
$\mathcal{U}\Gamma$-Poisson model	2.60	2.34	2.36

- **Effective duration** of the trial : **2.31 years**
  - The end of the trial was predicted with an error of **15 days, 10 mouths** before the expected date
  - **56 centres** would be enough for ending in 3 years.

- **Objectives :**
  - $N = 610$ patients
  - $T_R = 3$ years
  - $C_R = 77$ investigators centres
- **On-going studies :** after 1 year, after 1.5 year and after 2 years
- **The estimated duration** of the trial

The model	Time 1	Time 1.5	Time 2
Constant intensity	3.30	2.63	2.44
Γ-Poisson model	3.31	2.63	2.44
Π-Poisson model	2.63	2.39	2.36
$\mathcal{U}\Gamma$-Poisson model	2.60	2.34	2.36

- **Effective duration** of the trial : **2.31 years**
  - The end of the trial was predicted with an error of **15 days, 10 mouths** before the expected date
  - **56 centres** would be enough for ending in 3 years.

- **Objectives :**
  - $N = 610$ patients
  - $T_R = 3$ years
  - $C_R = 77$ investigators centres
- **On-going studies :** after 1 year, after 1.5 year and after 2 years
- **The estimated duration** of the trial

The model	Time 1	Time 1.5	Time 2
Constant intensity	3.30	2.63	2.44
Γ-Poisson model	3.31	2.63	2.44
Π-Poisson model	2.63	2.39	2.36
$\mathcal{U}\Gamma$-Poisson model	2.60	2.34	2.36

- **Effective duration** of the trial : **2.31 years**
  - The end of the trial was predicted with an error of **15 days, 10 mouths** before the expected date
  - **56 centres** would be enough for ending in 3 years.

Real case study : n=629 patients, N=91 centres

Data : $\vec{v} = (v(0), v(1), v(2), ...)$ where $v(j)$ is number of sites recruited j patients.

$$\vec{v} = (7, 11, 8, 8, 9, 8, 9, 7, 2, 4, 1, 3, 3, 4, 0, 0, 2, 1, 1, 2, 1, 0, 0, ..)$$

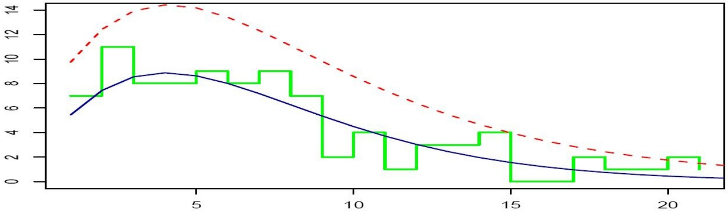
- **Real data :** step-wise green line
- **Fitted mean number of sites recruited j pts (theoretical) :** solid blue line
- **the mean + 2sd :** dashed red line

Huge variation among sites, rates are modelled using a gamma distribution and **fits real data**

Real case study : n=629 patients, N=91 centres

Data : $\vec{v} = (v(0), v(1), v(2), ...)$ where $v(j)$ is number of sites recruited j patients.

$$\vec{v} = (7, 11, 8, 8, 9, 8, 9, 7, 2, 4, 1, 3, 3, 4, 0, 0, 2, 1, 1, 2, 1, 0, 0, ..)$$



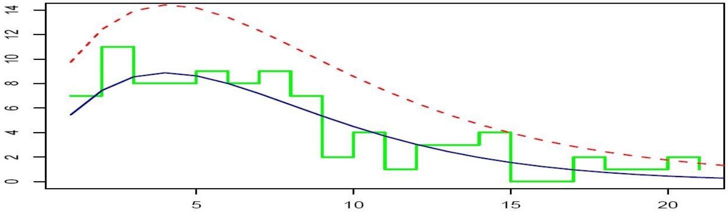
- **Real data :** step-wise green line
- **Fitted mean number of sites recruited *j* pts (theoretical) :** solid blue line
- **the mean + 2sd :** dashed red line

Huge variation among sites, rates are modelled using a gamma distribution and **fits real data**

Real case study : n=629 patients, N=91 centres

Data : $\vec{v} = (v(0), v(1), v(2), ...)$ where $v(j)$ is number of sites recruited j patients.

$$\vec{v} = (7, 11, 8, 8, 9, 8, 9, 7, 2, 4, 1, 3, 3, 4, 0, 0, 2, 1, 1, 2, 1, 0, 0, ..)$$



- **Real data :** step-wise green line
- **Fitted mean number of sites recruited *j* pts (theoretical) :** solid blue line
- **the mean + 2sd :** dashed red line

Huge variation among sites, rates are modelled using a gamma distribution and **fits real data**

Models investigated in (Anisimov, Mijoule, **Savy** (in progress))

- **Drop-out at the inclusion**

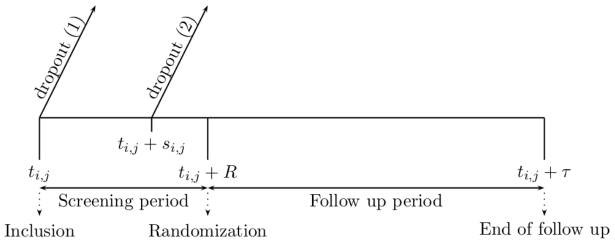
  modelled by a probability p_i in centre i

  $(p_1, \ldots, p_C)$ sample having a beta distribution
- **Drop-out during the screening period**

  modelled $s_{i,j}$ modelled by an exponential distribution of intensity θ_i

  $(\theta_1, \ldots, \theta_C)$ sample having a gamma distribution

Models investigated in (Anisimov, Mijoule, **Savy** (in progress))



- **Drop-out at the inclusion**

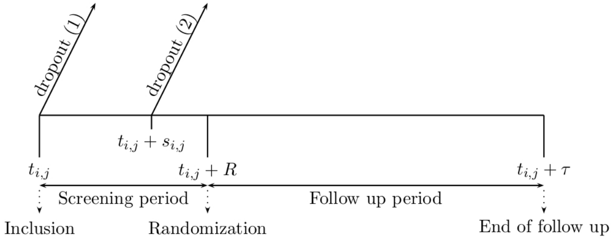
  modelled by a probability p_i in centre i

  $(p_1, \dots, p_C)$ sample having a beta distribution
- **Drop-out during the screening period**

  modelled $s_{i,j}$ modelled by an exponential distribution of intensity θ_i

  $(\theta_1, \dots, \theta_C)$ sample having a gamma distribution

Models investigated in (Anisimov, Mijoule, **Savy** (in progress))



- **Drop-out at the inclusion**

  modelled by a probability p_i in centre i

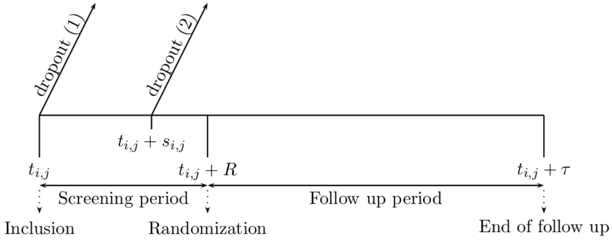
  $(p_1, \ldots, p_C)$ sample having a beta distribution

- **Drop-out during the screening period**

  modelled $s_{i,j}$ modelled by an exponential distribution of intensity θ_i

  $(\theta_1, \ldots, \theta_C)$ sample having a gamma distribution

Models investigated in (Anisimov, Mijoule, **Savy** (in progress))



- **Drop-out at the inclusion**

  modelled by a probability p_i in centre i

  $(p_1, \dots, p_C)$ sample having a beta distribution
- **Drop-out during the screening period**

  modelled $s_{i,j}$ modelled by an exponential distribution of intensity θ_i

  $(\theta_1, \dots, \theta_C)$ sample having a gamma distribution

- The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson.
- Drop-out process is directed by p a constant or $B(\psi_1, \psi_2)$.
- T_1 is an interim time.
  - τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
  - n_i number of **recruited patients** for centre i up to T_1
  - r_i number of **randomized patients** for centre i up to T_1

- The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson.
- Drop-out process is directed by p a constant or $B(\psi_1, \psi_2)$.
- T_1 is an interim time.
  - τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
  - n_i number of **recruited patients** for centre i up to T_1
  - r_i number of **randomized patients** for centre i up to T_1

- The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson.
- Drop-out process is directed by p a constant or $B(\psi_1, \psi_2)$.
- T_1 is an interim time.
  - τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
  - n_i number of **recruited patients** for centre i up to T_1
  - r_i number of **randomized patients** for centre i up to T_1

Theorem ((Anisimov, Mijoule, **Savy** (in progress)))

Given data $\{(n_i, r_i, \tau_i), 1 \leq i \leq C\}$, the log-likelihood function writes :

$$\mathcal{L}_1(\alpha, \beta, p) = \mathcal{L}_{1,1}(\alpha, \beta) + \mathcal{L}_{1,2}(p)$$

- *Notice the separation of the log-likelihood function (processes independent)*
- *$\mathcal{L}_{1,1}$ and $\mathcal{L}_{2,2}$ are explicit functions allowing optimisation.*

- The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson.
- Drop-out process is directed by p a constant or $B(\psi_1, \psi_2)$.
- T_1 is an interim time.
  - τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
  - n_i number of **recruited patients** for centre i up to T_1
  - r_i number of **randomized patients** for centre i up to T_1

Theorem ((Anisimov, Mijoule, **Savy** (in progress)))

Given data $\{(n_i, r_i, \tau_i), 1 \leq i \leq C\}$, the log-likelihood function writes :

$$\mathcal{L}_1(\alpha, \beta, \psi_1, \psi_2) = \mathcal{L}_{1,1}(\alpha, \beta) + \mathcal{L}_{1,2}(\psi_1, \psi_2)$$

- *Notice the separation of the log-likelihood function (processes independent)*
- *$\mathcal{L}_{1,1}$ and $\mathcal{L}_{2,2}$ are explicit functions allowing optimisation.*

- The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson.
- Drop-out process is directed by p a constant or $B(\psi_1, \psi_2)$.
- T_1 is an interim time.
  - τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
  - n_i number of **recruited patients** for centre i up to T_1
  - r_i number of **randomized patients** for centre i up to T_1

Theorem ((Anisimov, Mijoule, **Savy** (in progress)))

Given data $\{(n_i, r_i, \tau_i), 1 \leq i \leq C\}$, the log-likelihood function writes :

$$\mathcal{L}_1(\alpha, \beta, \psi_1, \psi_2) = \mathcal{L}_{1,1}(\alpha, \beta) + \mathcal{L}_{1,2}(\psi_1, \psi_2)$$

- *Notice the separation of the log-likelihood function (processes independent)*
- *$\mathcal{L}_{1,1}$ and $\mathcal{L}_{2,2}$ are explicit functions allowing optimisation.*

- The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson.
- Drop-out process is directed by p a constant or $B(\psi_1, \psi_2)$.
- T_1 is an interim time.
  - τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
  - n_i number of **recruited patients** for centre i up to T_1
  - r_i number of **randomized patients** for centre i up to T_1

Theorem ((Anisimov, Mijoule, **Savy** (in progress)))

Given data $\{(n_i, r_i, \tau_i), 1 \leq i \leq C\}$, the log-likelihood function writes :

$$\mathcal{L}_1(\alpha, \beta, \psi_1, \psi_2) = \mathcal{L}_{1,1}(\alpha, \beta) + \mathcal{L}_{1,2}(\psi_1, \psi_2)$$

- *Notice the separation of the log-likelihood function (processes independent)*
- *$\mathcal{L}_{1,1}$ and $\mathcal{L}_{2,2}$ are explicit functions allowing optimisation.*

The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson. T_1 is an interim time.

- τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
- n_i number of **recruited patients** for centre i up to T_1
- r_i number of **randomized patients** for centre i up to T_1
- ν_i number of **patients entered in screening period** for centre i in the interval $[T_1 - R, T_1]$

Theorem ((Anisimov, Mijoule, **Savy** (in progress)))

Given data $\{(n_i, r_i, \tau_i, \nu_i),\ 1 \leq i \leq C\}$, the predicted process of the number of randomized patients in centre i, $\{\widehat{\mathcal{R}}^i(t),\ t \geq T_1 + R\}$, expenses as

$$\widehat{\mathcal{R}}_i(t) = r_i + \mathrm{Bin}(\nu_i, \widehat{p}) + \Pi_{\widehat{p}\,\widehat{\lambda}_i}(t - T_1 - R).$$

$$\widehat{p} = \Big(\sum_{i=1}^{C} n_i\Big)^{-1} \sum_{i=1}^{C} r_i \quad \textit{and} \quad \widehat{\lambda}_i = \mathrm{Ga}(\widehat{\alpha} + n_i, \widehat{\beta} + \tau_i)$$

The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson. T_1 is an interim time.

- τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
- n_i number of **recruited patients** for centre i up to T_1
- r_i number of **randomized patients** for centre i up to T_1
- ν_i number of **patients entered in screening period** for centre i in the interval $[T_1 - R, T_1]$

Theorem ((Anisimov, Mijoule, **Savy** (in progress)))

Given data $\{(n_i, r_i, \tau_i, \nu_i), 1 \leq i \leq C\}$, the predicted process of the number of randomized patients in centre i, $\{\widehat{\mathcal{R}}^i(t),\ t \geq T_1 + R\}$, expenses as

$$\widehat{\mathcal{R}}_i(t) = r_i + \mathrm{Bin}(\nu_i, \widehat{p}_i) + \Pi_{\widehat{p}_i\,\widehat{\lambda}_i}(t - T_1 - R).$$

$$\widehat{p}_i = \mathrm{Beta}(\widehat{\psi}_1 + k_i, \widehat{\psi}_2 + n_i - k_i), \quad \textit{and} \quad \widehat{\lambda}_i = \mathrm{Ga}(\widehat{\alpha} + n_i, \widehat{\beta} + \tau_i)$$

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i
  - $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**
  - $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$

- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i
  - $\Rightarrow$ modelled by a **PP($(1-p_i)\lambda_i$)**
  - $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$
  - $\Rightarrow$ cost depend of the duration of the follow-up : $\sum_{0\leq T^i_j\leq t} g_i(t, T^i_j)$
    - g_i is a triangular function $g_i(t,s)=0$ when $t\leq s$
    - T^i_j are randomization time of the patient j by centre i

Improvement of patient recruitment' models

An additive model for cost

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i

  $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$

- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i

  $\Rightarrow$ modelled by a **PP($(1-p_i)\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$

  $\Rightarrow$ cost depend of the duration of the follow-up : $\sum_{0\leq T_j^i\leq t} g_i(t, T_j^i)$

  - g_i is a triangular function $g_i(t,s) = 0$ when $t \leq s$
  - T_j^i are randomization time of the patient j by centre i

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i

  $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$

- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i

  $\Rightarrow$ modelled by a **PP($(1-p_i)\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$

  $\Rightarrow$ cost depend of the duration of the follow-up : $\sum_{0\leq T_j^i\leq t} g_i(t, T_j^i)$

  - g_i is a triangular function $g_i(t,s) = 0$ when $t \leq s$
  - T_j^i are randomization time of the patient j by centre i

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i
  - $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**
  - $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$

- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i
  - $\Rightarrow$ modelled by a **PP($(1-p_i)\lambda_i$)**
  - $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$
  - $\Rightarrow$ cost depend of the duration of the follow-up : $\sum_{0\leq T_j^i\leq t} g_i(t, T_j^i)$
    - g_i is a triangular function $g_i(t,s)=0$ when $t\leq s$
    - T_j^i are randomization time of the patient j by centre i

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i

  $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$

- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i

  $\Rightarrow$ modelled by a **PP($(1-p_i)\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$

  $\Rightarrow$ cost depend of the duration of the follow-up : $\sum_{0\leq T_j^i\leq t} g_i(t, T_j^i)$

  - g_i is a triangular function $g_i(t,s) = 0$ when $t \leq s$
  - T_j^i are randomization time of the patient j by centre i

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i

  $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$

- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i

  $\Rightarrow$ modelled by a **PP($(1-p_i)\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$

  $\Rightarrow$ cost depend of the duration of the follow-up : $\sum_{0\leq T_j^i\leq t} g_i(t, T_j^i)$

  - g_i is a triangular function $g_i(t,s) = 0$ when $t \leq s$
  - T_j^i are randomization time of the patient j by centre i

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i

  $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**
  $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$

- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i

  $\Rightarrow$ modelled by a **PP($(1-p_i)\lambda_i$)**
  $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$
  $\Rightarrow$ cost depend of the duration of the follow-up : $\sum_{0 \leq T_j^i \leq t} g_i(t, T_j^i)$

  - g_i is a triangular function $g_i(t, s) = 0$ when $t \leq s$
  - T_j^i are randomization time of the patient j by centre i

In (Mijoule, Minois, Anisimov, **Savy** (forthcoming 2014)) ones considers the additive model for the cost generated by centre i :

$$\mathcal{C}_i(t) = J_i \mathcal{F}_i(t) + K_i \mathcal{R}_i(t) + \sum_{0 \leq T_j^i \leq t} g_i(t, T_j^i) + \underbrace{F_i + G_i\, t}_{\text{independent of patients}}$$

- The duration of the trial is the stopping time

$$T(N) = \inf_{t \geq 0} \{\mathcal{R}(t) \geq N\}$$

- The total cost of the trial is thus $\mathcal{C}(T(N)) = \sum_{i=1}^{C} \mathcal{C}_i(T(N))$
- In order to compute $\mathcal{C} = \mathbb{E}\left[\mathcal{C}(T(N))\right]$ we have to compute

$$\mathbb{E}\left[\int_0^{T(N)} g_i(T(N), s)\, d\mathcal{R}_i(s)\right].$$

It is not possible to use martingale arguments to compute such an expression

In (Mijoule, Minois, Anisimov, **Savy** (forthcoming 2014)) ones considers the additive model for the cost generated by centre i :

$$\mathcal{C}_i(t) = J_i\mathcal{F}_i(t) + K_i\mathcal{R}_i(t) + \int_0^t g_i(t,s)d\mathcal{R}_i(s) + \underbrace{F_i + G_i\,t}_{\text{independent of patients}}$$

- The duration of the trial is the stopping time

$$T(N) = \inf_{t \geq 0}\{\mathcal{R}(t) \geq N\}$$

- The total cost of the trial is thus $\mathcal{C}(T(N)) = \sum_{i=1}^{C} \mathcal{C}_i(T(N))$
- In order to compute $\mathcal{C} = \mathbb{E}\left[\mathcal{C}(T(N))\right]$ we have to compute

$$\mathbb{E}\left[\int_0^{T(N)} g_i(T(N),s)d\mathcal{R}_i(s)\right].$$

It is not possible to use martingale arguments to compute such an expression

In (Mijoule, Minois, Anisimov, **Savy** (forthcoming 2014)) ones considers the additive model for the cost generated by centre i :

$$\mathcal{C}_i(t) = J_i\mathcal{F}_i(t) + K_i\mathcal{R}_i(t) + \int_0^t g_i(t,s)d\mathcal{R}_i(s) + \underbrace{F_i + G_i\,t}_{\text{independent of patients}}$$

- The duration of the trial is the stopping time

$$T(N) = \inf_{t\geq 0}\{\mathcal{R}(t) \geq N\}$$

- The total cost of the trial is thus $\mathcal{C}(T(N)) = \sum_{i=1}^{C} \mathcal{C}_i(T(N))$
- In order to compute $\mathcal{C} = \mathbb{E}\left[\mathcal{C}(T(N))\right]$ we have to compute

$$\mathbb{E}\left[\int_0^{T(N)} g_i(T(N),s)d\mathcal{R}_i(s)\right].$$

It is not possible to use martingale arguments to compute such an expression

In (Mijoule, Minois, Anisimov, **Savy** (forthcoming 2014)) ones considers the additive model for the cost generated by centre i :

$$\mathcal{C}_i(t) = J_i \mathcal{F}_i(t) + K_i \mathcal{R}_i(t) + \int_0^t g_i(t,s) d\mathcal{R}_i(s) + \underbrace{F_i + G_i\, t}_{\text{independent of patients}}$$

- The duration of the trial is the stopping time

$$T(N) = \inf_{t \geq 0} \{\mathcal{R}(t) \geq N\}$$

- The total cost of the trial is thus $\mathcal{C}(T(N)) = \sum_{i=1}^{C} \mathcal{C}_i(T(N))$
- In order to compute $\mathcal{C} = \mathbb{E}\left[\mathcal{C}(T(N))\right]$ we have to compute

$$\mathbb{E}\left[\int_0^{T(N)} g_i(T(N), s) d\mathcal{R}_i(s)\right].$$

It is not possible to use martingale arguments to compute such an expression

In (Mijoule, Minois, Anisimov, **Savy** (forthcoming 2014)) ones considers the additive model for the cost generated by centre i :

$$\mathcal{C}_i(t) = J_i\mathcal{F}_i(t) + K_i\mathcal{R}_i(t) + \int_0^t g_i(t,s)d\mathcal{R}_i(s) + \underbrace{F_i + G_i\,t}_{\text{independent of patients}}$$

- The duration of the trial is the stopping time

$$T(N) = \inf_{t\geq 0}\{\mathcal{R}(t) \geq N\}$$

- The total cost of the trial is thus $\mathcal{C}(T(N)) = \sum_{i=1}^{C}\mathcal{C}_i(T(N))$
- In order to compute $\mathcal{C} = \mathbb{E}\left[\mathcal{C}(T(N))\right]$ we have to compute

$$\mathbb{E}\left[\int_0^{T(N)} g_i(T(N),s)d\mathcal{R}_i(s)\right].$$

It is not possible to use martingale arguments to compute such an expression

In (Mijoule, Minois, Anisimov, **Savy** (forthcoming 2014)) ones considers the additive model for the cost generated by centre i :

$$\mathcal{C}_i(t) = J_i \mathcal{F}_i(t) + K_i \mathcal{R}_i(t) + \int_0^t g_i(t,s) d\mathcal{R}_i(s) + \underbrace{F_i + G_i\, t}_{\text{independent of patients}}$$

- The duration of the trial is the stopping time

$$T(N) = \inf_{t \geq 0} \{\mathcal{R}(t) \geq N\}$$

- The total cost of the trial is thus $\mathcal{C}(T(N)) = \sum_{i=1}^{C} \mathcal{C}_i(T(N))$
- In order to compute $\mathcal{C} = \mathbb{E}\left[\mathcal{C}(T(N))\right]$ we have to compute

$$\mathbb{E}\left[\int_0^{T(N)} g_i(T(N), s) d\mathcal{R}_i(s)\right].$$

It is not possible to use martingale arguments to compute such an expression

Theorem ((Mijoule, Minois, Anisimov, **Savy** (2014)))

- *Assume* $(\lambda_i)_{1\leq i\leq C}$ *and* $(p_i)_{1\leq i\leq C}$ ***are known***

$\Longrightarrow$ ***we have an explicit expression of*** $\mathcal{C}$

- *Assume* $\lambda_i \sim \Gamma(\alpha,\beta)$ *and* $p_i \sim B(\psi_1,\psi_2)$
- *Consider an interim time* T_1, *and consider that the i-th centre has*
    - *screened* n_i *patients*
    - *randomized* r_i *patients*
- *Given* (n_i, r_i) *the posterior distribution of*
    - *the rate is* $\lambda_i \sim \Gamma(\alpha+n_i, \beta+T_1)$
    - *the probability of screening failure is* $p_i \sim B(\psi_1+r_i, \psi_2+n_i-r_i)$

$\Longrightarrow$ ***we can compute*** $\mathcal{C}$ ***by means of Monte Carlo simulation***

INSTITUT de MATHEMATIQUES de TOULOUSE

Theorem ((Mijoule, Minois, Anisimov, **Savy** (2014)))

- *Assume $(\lambda_i)_{1\leq i\leq C}$ and $(p_i)_{1\leq i\leq C}$* ***are known***

$\Longrightarrow$ ***we have an explicit expression of $\mathcal{C}$***

- *Assume $\lambda_i \sim \Gamma(\alpha, \beta)$ and $p_i \sim B(\psi_1, \psi_2)$*
- *Consider an interim time T_1, and consider that the i-th centre has*
  - *screened n_i patients*
  - *randomized r_i patients*
- *Given (n_i, r_i) the posterior distribution of*
  - *the rate is $\lambda_i \sim \Gamma(\alpha + n_i, \beta + T_1)$*
  - *the probability of screening failure is $p_i \sim B(\psi_1 + r_i, \psi_2 + n_i - r_i)$*

$\Longrightarrow$ ***we can compute $\mathcal{C}$ by means of Monte Carlo simulation***

Theorem ((Mijoule, Minois, Anisimov, **Savy** (2014)))

- *Assume* $(\lambda_i)_{1\leq i\leq C}$ *and* $(p_i)_{1\leq i\leq C}$ ***are known***

$\Longrightarrow$ ***we have an explicit expression of*** $\mathcal{C}$

- *Assume* $\lambda_i \sim \Gamma(\alpha, \beta)$ *and* $p_i \sim B(\psi_1, \psi_2)$
- *Consider an interim time* T_1*, and consider that the i-th centre has*
    - *screened* n_i *patients*
    - *randomized* r_i *patients*
- *Given* (n_i, r_i) *the posterior distribution of*
    - *the rate is* $\lambda_i \sim \Gamma(\alpha + n_i, \beta + T_1)$
    - *the probability of screening failure is* $p_i \sim B(\psi_1 + r_i, \psi_2 + n_i - r_i)$

$\Longrightarrow$ ***we can compute*** $\mathcal{C}$ ***by means of Monte Carlo simulation***

Theorem ((Mijoule, Minois, Anisimov, **Savy** (2014)))

- *Assume* $(\lambda_i)_{1\leq i\leq C}$ *and* $(p_i)_{1\leq i\leq C}$ ***are known***

  $\Longrightarrow$ ***we have an explicit expression of*** $\mathcal{C}$

- *Assume* $\lambda_i \sim \Gamma(\alpha, \beta)$ *and* $p_i \sim B(\psi_1, \psi_2)$
- *Consider an interim time* T_1*, and consider that the* i*-th centre has*
  - *screened* n_i *patients*
  - *randomized* r_i *patients*
- *Given* (n_i, r_i) *the posterior distribution of*
  - *the rate is* $\lambda_i \sim \Gamma(\alpha + n_i, \beta + T_1)$
  - *the probability of screening failure is* $p_i \sim B(\psi_1 + r_i, \psi_2 + n_i - r_i)$

  $\Longrightarrow$ ***we can compute*** $\mathcal{C}$ ***by means of Monte Carlo simulation***

Theorem ((Mijoule, Minois, Anisimov, **Savy** (2014)))

- *Assume* $(\lambda_i)_{1\leq i\leq C}$ *and* $(p_i)_{1\leq i\leq C}$ ***are known***

  $\Longrightarrow$ ***we have an explicit expression of*** $\mathcal{C}$

- *Assume* $\lambda_i \sim \Gamma(\alpha, \beta)$ *and* $p_i \sim B(\psi_1, \psi_2)$
- *Consider an interim time* T_1*, and consider that the* i*-th centre has*
  - *screened* n_i *patients*
  - *randomized* r_i *patients*
- *Given* (n_i, r_i) *the posterior distribution of*
  - *the rate is* $\lambda_i \sim \Gamma(\alpha + n_i, \beta + T_1)$
  - *the probability of screening failure is* $p_i \sim B(\psi_1 + r_i, \psi_2 + n_i - r_i)$

$\Longrightarrow$ ***we can compute*** $\mathcal{C}$ ***by means of Monte Carlo simulation***

Theorem ((Mijoule, Minois, Anisimov, **Savy** (2014)))

- *Assume $(\lambda_i)_{1\leq i\leq C}$ and $(p_i)_{1\leq i\leq C}$* ***are known***

  $\Longrightarrow$ ***we have an explicit expression of $\mathcal{C}$***

- *Assume $\lambda_i \sim \Gamma(\alpha,\beta)$ and $p_i \sim B(\psi_1,\psi_2)$*
- *Consider an interim time T_1, and consider that the i-th centre has*
  - *screened n_i patients*
  - *randomized r_i patients*
- *Given (n_i, r_i) the posterior distribution of*
  - *the rate is $\lambda_i \sim \Gamma(\alpha + n_i, \beta + T_1)$*
  - *the probability of screening failure is $p_i \sim B(\psi_1 + r_i, \psi_2 + n_i - r_i)$*

  $\Longrightarrow$ ***we can compute $\mathcal{C}$ by means of Monte Carlo simulation***

INSTITUT de MATHEMATIQUES de TOULOUSE

Assume the closure of centre j, denote

- $T^j(N)$ the duration of the trial without centre j
- $\mathcal{C}^j(t)$ the cost of the trial at time t without centre j

- By means of Monte Carlo simulation we are able to evaluate the variation of cost due to centre j closure :

$$\Delta\mathcal{C}_j = \mathbb{E}\left[\mathcal{C}(T(N)) - \mathcal{C}^j(T^j(N))\right]$$

- Consider $(\Delta\mathcal{C}_j, T^j(N))$ to decide on the closure of centre j.

Assume the closure of centre j, denote

- $T^j(N)$ the duration of the trial without centre j
- $\mathcal{C}^j(t)$ the cost of the trial at time t without centre j

- By means of Monte Carlo simulation we are able to evaluate the variation of cost due to centre j closure :

$$\Delta\mathcal{C}_j = \mathbb{E}\left[\mathcal{C}(T(N)) - \mathcal{C}^j(T^j(N))\right]$$

- Consider $(\Delta\mathcal{C}_j, T^j(N))$ to decide on the closure of centre j.

Assume the closure of centre j, denote

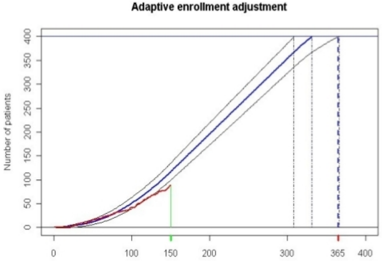
- $T^j(N)$ the duration of the trial without centre j
- $\mathcal{C}^j(t)$ the cost of the trial at time t without centre j

- By means of Monte Carlo simulation we are able to evaluate the variation of cost due to centre j closure :

$$\Delta\mathcal{C}_j = \mathbb{E}\left[\mathcal{C}(T(N)) - \mathcal{C}^j(T^j(N))\right]$$

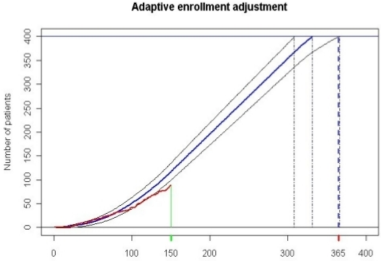
- Consider $(\Delta\mathcal{C}_j, T^j(N))$ to decide on the closure of centre j.

Study design : Sites : 70, patient's target : 400, enrolment duration :1 year
Sites initiated in 5-month period,
half of sites will be closed in two months before the end of enrolment



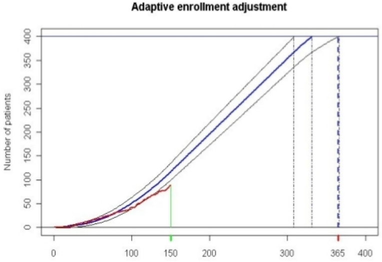
- **Initial design :** to complete with 90% confidence.
  Predictive area : mean and confidence bounds.
- **Interim analysis after 150 days :** 88 pts recruited.
  Real enrolment is slower than predicted.
- **Interim adjustment :**
  to complete with 90% confidence : 22 new sites to add

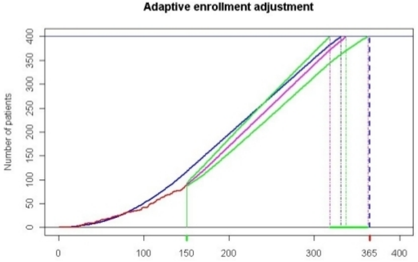
Study design : Sites : 70, patient's target : 400, enrolment duration :1 year
Sites initiated in 5-month period,
half of sites will be closed in two months before the end of enrolment



- **Initial design :** to complete with 90% confidence.
  Predictive area : mean and confidence bounds.
- **Interim analysis after 150 days :** 88 pts recruited.
  Real enrolment is slower than predicted.
- **Interim adjustment :**
  to complete with 90% confidence : 22 new sites to add

Study design : Sites : 70, patient's target : 400, enrolment duration :1 year
Sites initiated in 5-month period,
half of sites will be closed in two months before the end of enrolment





- **Initial design :** to complete with 90% confidence.
  Predictive area : mean and confidence bounds.
- **Interim analysis after 150 days :** 88 pts recruited.
  Real enrolment is slower than predicted.
- **Interim adjustment :**
  to complete with 90% confidence : 22 new sites to add

Modelling enrolment and **hierarchic follow-up** processes is a basic methodology for **forecasting future performance** and developing different **triggers** :

- **Triggers for detecting outliers :**

  Late-start, inactive, high number of AE, low-enrolling, etc.

- **Predictive triggers (interim time analysis, data-driven) :**

  Predicting future behavior and alarm unusual sites

  Create dynamic forecasts in future time intervals

  Opportunities for optimal decision-making (sites, costs, risks).

Current triggers for RBM usually use assumptions of **normality** and detect unusual behaviour within cohort using Mean and SD :

$$X > Mean(cohort) + K * SD(cohort), K = 1, 2, 3$$

Many of variables describing trial operation are **rather far from the normality**

Modelling enrolment and **hierarchic follow-up** processes is a basic methodology for **forecasting future performance** and developing different **triggers** :

- **Triggers for detecting outliers :**

  Late-start, inactive, high number of AE, low-enrolling, etc.

- **Predictive triggers (interim time analysis, data-driven) :**

  Predicting future behavior and alarm unusual sites

  Create dynamic forecasts in future time intervals

  Opportunities for optimal decision-making (sites, costs, risks).

Current triggers for RBM usually use assumptions of **normality** and detect unusual behaviour within cohort using Mean and SD :

$$X > Mean(cohort) + K * SD(cohort), K = 1, 2, 3$$

Many of variables describing trial operation are **rather far from the normality**

Modelling enrolment and **hierarchic follow-up** processes is a basic methodology for **forecasting future performance** and developing different **triggers** :

- **Triggers for detecting outliers :**

  Late-start, inactive, high number of AE, low-enrolling, etc.

- **Predictive triggers (interim time analysis, data-driven) :**

  Predicting future behavior and alarm unusual sites

  Create dynamic forecasts in future time intervals

  Opportunities for optimal decision-making (sites, costs, risks).

Current triggers for RBM usually use assumptions of **normality** and detect unusual behaviour within cohort using Mean and SD :

$$X > Mean(cohort) + K * SD(cohort), K = 1, 2, 3$$

Many of variables describing trial operation are **rather far from the normality**

Modelling enrolment and **hierarchic follow-up** processes is a basic methodology for **forecasting future performance** and developing different **triggers** :

- **Triggers for detecting outliers :**

  Late-start, inactive, high number of AE, low-enrolling, etc.

- **Predictive triggers (interim time analysis, data-driven) :**

  Predicting future behavior and alarm unusual sites

  Create dynamic forecasts in future time intervals

  Opportunities for optimal decision-making (sites, costs, risks).

Current triggers for RBM usually use assumptions of **normality** and detect unusual behaviour within cohort using Mean and SD :

$$X > Mean(cohort) + K * SD(cohort), K = 1, 2, 3$$

Many of variables describing trial operation are **rather far from the normality**

Modelling enrolment and **hierarchic follow-up** processes is a basic methodology for **forecasting future performance** and developing different **triggers** :

- **Triggers for detecting outliers :**

  Late-start, inactive, high number of AE, low-enrolling, etc.

- **Predictive triggers (interim time analysis, data-driven) :**

  Predicting future behavior and alarm unusual sites

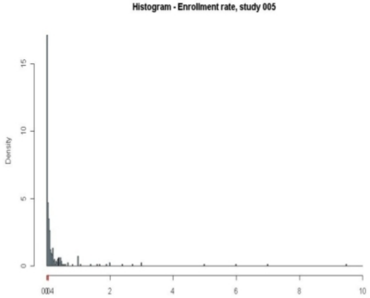
  Create dynamic forecasts in future time intervals

  Opportunities for optimal decision-making (sites, costs, risks).

Current triggers for RBM usually use assumptions of **normality** and detect unusual behaviour within cohort using Mean and SD :

$$X > Mean(cohort) + K * SD(cohort), K = 1, 2, 3$$

Many of variables describing trial operation are **rather far from the normality**

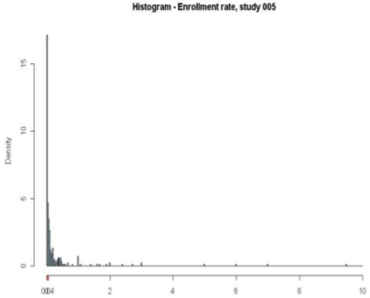


Histogram of the enrolment rate
(# of patients)/(site enrolment duration)

- far from normal distribution
- heavy tailed
- Adequate model : Poisson mixed with gamma

Histogram of time from Last Patient Enrolled till current time

- far from normal distribution
- heavy tailed
- Adequate model : Exponential mixed with gamma or Pareto



Histogram of the enrolment rate

(# of patients)/(site enrolment duration)

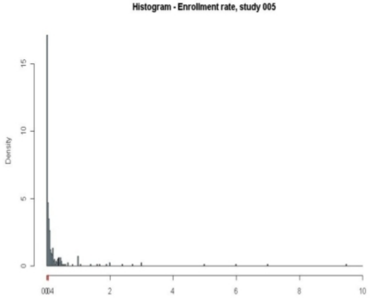
- far from normal distribution
- heavy tailed
- Adequate model : Poisson mixed with gamma

Histogram of time from Last Patient Enrolled till current time

- far from normal distribution
- heavy tailed
- Adequate model : Exponential mixed with gamma or Pareto

Decisions tools

Site performance and risk-based monitoring (Vladimir Anisimov)



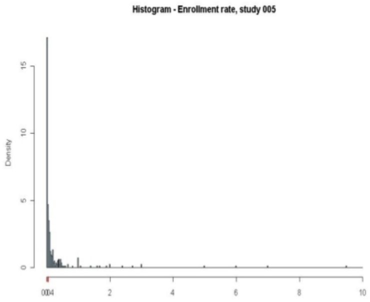
Histogram of the enrolment rate

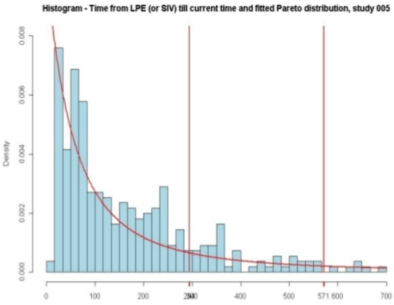
(# of patients)/(site enrolment duration)

- far from normal distribution
- heavy tailed
- Adequate model : Poisson mixed with gamma

Histogram of time from Last Patient Enrolled till current time

- far from normal distribution
- heavy tailed
- Adequate model : Exponential mixed with gamma or Pareto





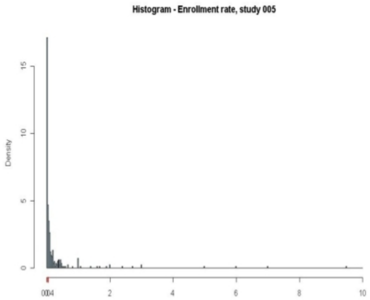
Histogram of the enrolment rate

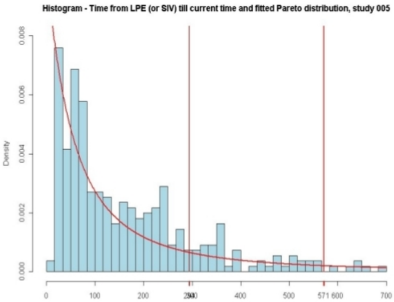
(# of patients)/(site enrolment duration)

- far from normal distribution
- heavy tailed
- Adequate model : Poisson mixed with gamma

Histogram of time from Last Patient Enrolled till current time

- far from normal distribution
- heavy tailed
- Adequate model : Exponential mixed with gamma or Pareto





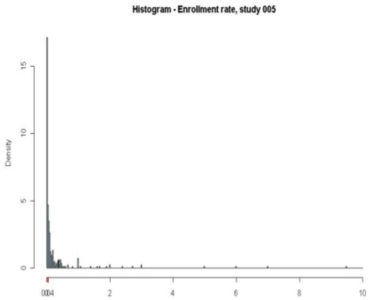
Histogram of the enrolment rate

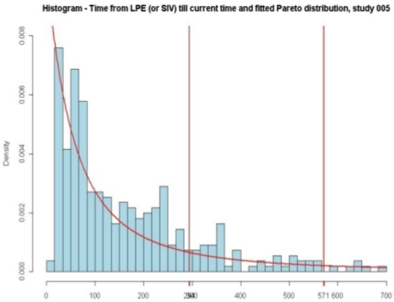
(# of patients)/(site enrolment duration)

- far from normal distribution
- heavy tailed
- Adequate model : Poisson mixed with gamma

Histogram of time from Last Patient Enrolled till current time

- far from normal distribution
- heavy tailed
- Adequate model : Exponential mixed with gamma or Pareto





Histogram of the enrolment rate

(# of patients)/(site enrolment duration)

- far from normal distribution
- heavy tailed
- Adequate model : Poisson mixed with gamma

Histogram of time from Last Patient Enrolled till current time

- far from normal distribution
- heavy tailed
- Adequate model : Exponential mixed with gamma or Pareto

Interim analysis, real case study, 330 active sites.

SID	# of patients	Enrolment duration		
1004	1	166		
1006	0	268		
1007	2	533		
1009	1	190		
1011	0	124		
1012	1	595		
1013	3	450		
1014	5	488		
1017	0	494		
1022	3	486		
1029	0	5		
1201	2	424		
1203	2	316		
1901	25	180		
1904	3	347		
1905	5	550		
1906	10	534		

Poisson-Gamma model of enrolment + Data-driven Bayesian re-estimation of rates
$\Longrightarrow$ **Predictive probabilities** for the next 4-month period

- **enrol no patients**
- **Enrol at least one**

Indicator of **hight** blueblow performance

Interim analysis, real case study, 330 active sites.

SID	# of patients	Enrolment duration		
1004	1	166		
1006	0	268		
1007	2	533		
1009	1	190		
1011	0	124		
1012	1	595		
1013	3	450		
1014	5	488		
1017	0	494		
1022	3	486		
1029	0	5		
1201	2	424		
1203	2	316		
1901	25	180		
1904	3	347		
1905	5	550		
1906	10	534		

Poisson-Gamma model of enrolment + Data-driven Bayesian re-estimation of rates

$\Longrightarrow$ **Predictive probabilities** for the next 4-month period

- **enrol no patients**
- **Enrol at least one**

Indicator of **hight** blueblow performance

Interim analysis, real case study, 330 active sites.

SID	# of patients	Enrolment duration	Prob Zero	
1004	1	166	0.462	
1006	0	268	0.716	
1007	2	533	0.578	
1009	1	190	0.485	
1011	0	124	0.614	
1012	1	595	0.705	
1013	3	450	0.445	
1014	5	488	0.325	
1017	0	494	0.801	
1022	3	486	0.465	
1029	0	5	0.451	
1201	2	424	0.525	
1203	2	316	0.457	
1901	25	180	0.0002	
1904	3	347	0.381	
1905	5	550	0.358	
1906	10	534	0.151	

Poisson-Gamma model of enrolment + Data-driven Bayesian re-estimation of rates

$\Longrightarrow$ **Predictive probabilities** for the next 4-month period

- **enrol no patients**
- **Enrol at least one**

Indicator of **hight** blueblow performance

Interim analysis, real case study, 330 active sites.

SID	# of patients	Enrolment duration	Prob Zero	Prob <=1
1004	1	166	0.462	0.764
1006	0	268	0.716	0.926
1007	2	533	0.578	0.87
1009	1	190	0.485	0.785
1011	0	124	0.614	0.862
1012	1	595	0.705	0.933
1013	3	450	0.445	0.773
1014	5	488	0.325	0.66
1017	0	494	0.801	0.964
1022	3	486	0.465	0.791
1029	0	5	0.451	0.717
1201	2	424	0.525	0.832
1203	2	316	0.457	0.775
1901	25	180	0.0002	0.001
1904	3	347	0.381	0.71
1905	5	550	0.358	0.697
1906	10	534	0.151	0.413

Poisson-Gamma model of enrolment + Data-driven Bayesian re-estimation of rates

$\Longrightarrow$ **Predictive probabilities** for the next 4-month period

- **enrol no patients**
- **Enrol at least one**

Indicator of **hight** blueblow performance

Interim analysis, real case study, 330 active sites.

SID	# of patients	Enrolment duration	Prob Zero	Prob <=1
1004	1	166	0.462	0.764
1006	0	268	0.716	0.926
1007	2	533	0.578	0.87
1009	1	190	0.485	0.785
1011	0	124	0.614	0.862
1012	1	595	0.705	0.933
1013	3	450	0.445	0.773
1014	5	488	0.325	0.66
1017	0	494	0.801	0.964
1022	3	486	0.465	0.791
1029	0	5	0.451	0.717
1201	2	424	0.525	0.832
1203	2	316	0.457	0.775
1901	25	180	0.0002	0.001
1904	3	347	0.381	0.71
1905	5	550	0.358	0.697
1906	10	534	0.151	0.413

Poisson-Gamma model of enrolment + Data-driven Bayesian re-estimation of rates

$\Longrightarrow$ **Predictive probabilities** for the next 4-month period

- **enrol no patients**
- **Enrol at least one**

Indicator of **hight** blueblow performance

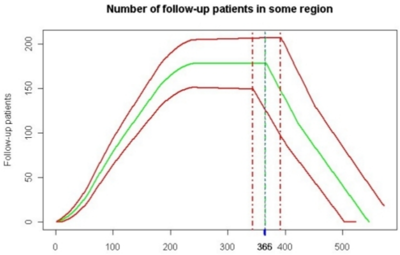
Study design : Sites : 200, patient's target : 800, enrolment duration :1 year
4 visits in total, each after 60 days, Follow-up period L=180 days

- **Predictive number of follow-up patients**

  Mean, Low and Upper 90% bounds for a Region with 100 sites
- **Predictive number of Visits No. 3**

  Mean, Low and Upper 90% bounds for a Region with 100 sites

Study design : Sites : 200, patient's target : 800, enrolment duration :1 year
4 visits in total, each after 60 days, Follow-up period L=180 days



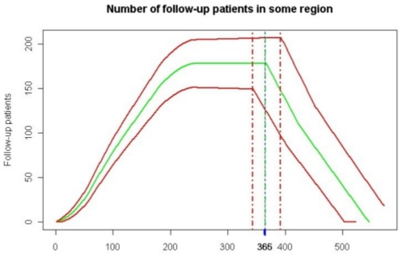
- **Predictive number of follow-up patients**

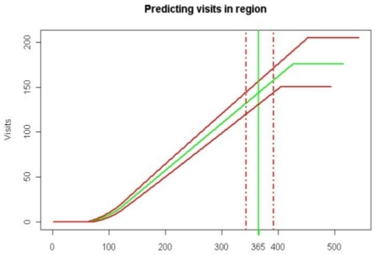
  Mean, Low and Upper 90% bounds for a Region with 100 sites

- **Predictive number of Visits No. 3**

  Mean, Low and Upper 90% bounds for a Region with 100 sites

Study design : Sites : 200, patient's target : 800, enrolment duration :1 year
4 visits in total, each after 60 days, Follow-up period L=180 days





- **Predictive number of follow-up patients**

  Mean, Low and Upper 90% bounds for a Region with 100 sites
- **Predictive number of Visits No. 3**

  Mean, Low and Upper 90% bounds for a Region with 100 sites

Thank you for your attention...

- **Nathan Minois**
  INSERM 1027 - Toulouse
- **Vladimir Anisimov**
  Quintiles - UK
- **Stéphanie Savy**
  INSERM 1027 - Toulouse
- **Guillaume Mijoule**
  Université Paris XI
- **Sandrine Andrieu**
  INSERM 1027 - Toulouse
- **Valérie Lauwers-Cances**
  CHU - Toulouse

This research has benefited from the help of IRESP during the call for proposals launched in 2012 in the setting of Cancer Plan 2009-2013

Thank you for your attention...

- **Nathan Minois**
  INSERM 1027 - Toulouse
- **Vladimir Anisimov**
  Quintiles - UK
- **Stéphanie Savy**
  INSERM 1027 - Toulouse
- **Guillaume Mijoule**
  Université Paris XI
- **Sandrine Andrieu**
  INSERM 1027 - Toulouse
- **Valérie Lauwers-Cances**
  CHU - Toulouse

This research has benefited from the help of IRESP during the call for proposals launched in 2012 in the setting of Cancer Plan 2009-2013

Thank you for your attention...

- **Nathan Minois**
  INSERM 1027 - Toulouse
- **Vladimir Anisimov**
  Quintiles - UK
- **Stéphanie Savy**
  INSERM 1027 - Toulouse
- **Guillaume Mijoule**
  Université Paris XI
- **Sandrine Andrieu**
  INSERM 1027 - Toulouse
- **Valérie Lauwers-Cances**
  CHU - Toulouse

This research has benefited from the help of IRESP during the call for proposals launched in 2012 in the setting of Cancer Plan 2009-2013