On the use of Fleming and Harrington test to detect Late Survival Differences in clinical trials

Valérie GARES\textsuperscript{1,3}

Sandrine ANDRIEU\textsuperscript{1,3}, Jean-François DUPUY\textsuperscript{2}

and Nicolas SAVY\textsuperscript{1,3}

\textsuperscript{1}INSERM Unit 1027 Ageing Team - Toulouse
\textsuperscript{2}INSA - Rennes and \textsuperscript{3}Paul Sabatier University - Toulouse

September 21, 2012
Consider a clinical trial with two groups:

1. patients who receive a drug
2. patients who receive a placebo

The data of the primary endpoint are time-to-event (dementia). The duration of follow-up is \( \tau \) years.
Consider a clinical trial with two groups:

1. patients who receive a drug
2. patients who receive a placebo

The data of the primary endpoint are time-to-event (dementia).

The duration of follow-up is $\tau$ years.
Consider a clinical trial with two groups:

1. patients who receive a drug
2. patients who receive a placebo

The data of the primary endpoint are time-to-event (dementia).
The duration of follow-up is $\tau$ years.
Consider a clinical trial with two groups:
1. patients who receive a drug
2. patients who receive a placebo

The data of the primary endpoint are time-to-event (dementia).
The duration of follow-up is $\tau$ years.

We test:

$H_0 : S_P(t) = S_T(t), \forall t,$

$H_1 : S_P(t) \neq S_T(t),$

where $S_P$ and $S_T$ are the survival functions for placebo and treated groups.
Consider a clinical trial with two groups:

- 1 patients who receive a drug
- 2 patients who receive a placebo

The data of the primary endpoint are time-to-event (dementia).
The duration of follow-up is $\tau$ years.

We test:

$\mathcal{H}_0 : \lambda_P(t) = \lambda_T(t), \forall t,$

$\mathcal{H}_1 : \lambda_P(t) \neq \lambda_T(t),$

where $\lambda_P$ and $\lambda_T$ are the hazard functions for placebo and treated groups.
Consider a clinical trial with two groups:

1. patients who receive a drug
2. patients who receive a placebo

The data of the primary endpoint are time-to-event (dementia). The duration of follow-up is $\tau$ years.

We test:

$H_0 : \lambda_P(t) = \lambda_T(t), \forall t,$

$H_1 : \lambda_P(t) \neq \lambda_T(t),$

where $\lambda_P$ and $\lambda_T$ are the hazard functions for placebo and treated groups.

Investigators usually use the Logrank test in the statistical design.
In prevention of Alzheimer disease,

- Some observational studies show some factors can protect against occurrence of dementia.

- No clinical trial has demonstrated a protective effect of these factors.
In prevention of Alzheimer disease,

- Some **observational studies show some factors can protect** against occurrence of dementia.
  - hormone treatments
  - antioxidant treatments
  - lifestyle factors (fish intake, physical exercise ...)
  - ...

- No clinical trial has demonstrated a protective effect of these factors.
In prevention of Alzheimer disease,

- Some **observational studies show some factors can protect** against occurrence of dementia.
  - *hormone treatments*
  - *antioxidant treatments*
  - *lifestyle factors (fish intake, physical exercise ...)*
  - ...

- **No clinical trial has demonstrated a protective effect of these factors.**
In prevention of Alzheimer disease,

- Some **observational studies show some factors can protect** against occurrence of dementia.
  - *hormone treatments*
  - *antioxidant treatments*
  - *lifestyle factors (fish intake, physical exercise ...)*
  - ...

- **No clinical trial has demonstrated a protective effect** of these factors.
Two possible assumptions:

1. No protective effect of these factors
2. There is an effect of these factors but the Logrank test is not able to capture this effect.
Problematic

Two possible assumptions:

1. **No protective effect** of these factors

2. **There is an effect** of these factors
   - **BUT** the *Logrank test* is not able to capture this effect.
     - Lack of power
     - The computation of the sample size avoids this option.
     - Assumptions not fulfilled or not optimal
Two possible assumptions:

1. No protective effect of these factors

2. There is an effect of these factors
   BUT the Logrank test is not able to capture this effect.

- Lack of power
  The computation of the sample size avoids this option.

- Assumptions not fulfilled or not optimal
Problematic

Two possible assumptions:

1. **No protective effect** of these factors

2. **There is an effect** of these factors
   
   **BUT** the *Logrank test* is not able to capture this effect.
   
   - Lack of power
     
     The computation of the sample size avoids this option.
   
   - Assumptions not fulfilled or not optimal
Two possible assumptions:

1. **No protective effect** of these factors

2. **There is an effect** of these factors
   - **BUT** the *Logrank test* is not able to capture this effect.
   - Lack of power
     - The computation of the sample size avoids this option.
   - Assumptions not fulfilled or not optimal
Logrank test

- It tests

\[ H_0 : \lambda_P(t) = \lambda_T(t), \forall t \]
\[ H_1 : \lambda_P(t) \neq \lambda_T(t). \]

- There are many ways to be \( H_1 \).
  BUT Logrank test is optimal (in a sense defined later) to detect

\[ H_0 : \lambda_P(t) = \lambda_T(t), \forall t \]
\[ H_1 : \lambda_P(t) = \text{Constant} \times \lambda_T(t) \forall t. \]

- Logrank test is thus optimal for proportional effects during the follow up period (Cox model).
Logrank test

- It tests

\[ \mathcal{H}_0 : \lambda_P(t) = \lambda_T(t), \forall t \]
\[ \mathcal{H}_1 : \lambda_P(t) \neq \lambda_T(t). \]

- There are many ways to be \( \mathcal{H}_1 \).

BUT Logrank test is optimal (in a sense defined later) to detect

\[ \mathcal{H}_0 : \lambda_P(t) = \lambda_T(t), \forall t \]
\[ \mathcal{H}_1 : \lambda_P(t) = \text{Constant} \times \lambda_T(t) \forall t. \]

- Logrank test is thus optimal for \textit{proportional effects} during the follow up period (Cox model).
Logrank test

- It tests

\[ \mathcal{H}_0 : \lambda_P(t) = \lambda_T(t), \forall t \]
\[ \mathcal{H}_1 : \lambda_P(t) \neq \lambda_T(t). \]

- There are many ways to be \( \mathcal{H}_1 \).
  
  BUT Logrank test is optimal (in a sense defined later) to detect

\[ \mathcal{H}_0 : \lambda_P(t) = \lambda_T(t), \forall t \]
\[ \mathcal{H}_1 : \lambda_P(t) = \text{Constant} \times \lambda_T(t) \forall t. \]

- Logrank test is thus optimal for **proportional effects** during the follow up period (Cox model).
**Prevention studies**

**Special feature of Prevention studies**

By the very definition of prevention studies we can observe Late Effects. The assumption of proportional effects is not realistic.

- $t^*$: Instant when we begin to observe an effect
- $\delta$: Difference between survival functions at the end point
Prevention studies

Special feature of Prevention studies

By the very definition of prevention studies we can observe **Late Effects**. The assumption of proportional effects is not realistic.

- $t^*$: Instant when we begin to observe an effect
- $\delta$: Difference between survival functions at the end point
Prevention studies

Special feature of Prevention studies

By the very definition of prevention studies we can observe \textbf{Late Effects}. The assumption of proportional effects is not realistic.

- $t^*$: Instant when we begin to observe an effect
- $\delta$: Difference between survival functions at the end point

\[ S_T \text{ (Proportional risks)} \]
\[ S_P \text{ (placebo)} \]
Prevention studies

Special feature of Prevention studies

By the very definition of prevention studies we can observe **Late Effects**. The assumption of proportional effects is not realistic.

\[ S_T \] (Late effect)

\[ S_T \] (Proportional risks)

\[ S_P \] (placebo)

- \( t^* \): Instant when we begin to observe an effect
- \( \delta \): Difference between survival functions at the end point
Prevention studies

Special feature of Prevention studies

By the very definition of prevention studies we can observe **Late Effects**. The assumption of proportional effects is not realistic.

- $S_T$ (Late effect)
- $S_T$ (Proportional risks)
- $S_P$ (placebo)

- $t^*$ : Instant when we begin to observe an effect
- $\delta$ : Difference between survival functions at the end point
Prevention studies

**Special feature of Prevention studies**

By the very definition of prevention studies we can observe **Late Effects**. The assumption of proportional effects is not realistic.

- $S_T$ (Late effect)
- $S_T$ (Proportional risks)
- $S_P$ (placebo)

$t^*$ : Instant when we begin to observe an effect

$\delta$ : Difference between survival functions at the end point
Many tests have been defined to detect late (early) effects. [Gehan(1965), Peto & Peto(1972), Tarone & Ware(1977), Prentice(1978), Harrington & Fleming(1982), Fleming & Harrington(1991)]

Our objective is to propose methodology for the use of Fleming and Harrington test in the setting of clinical trials.

The constraints of a Clinical trial design are:

1. The test has to be chosen before the knowledge of the data.
2. All the parameters have to be chosen by the investigator.
3. The sample size can (and has to) be calculated.
Many tests have been defined to detect late (early) effects. [Gehan(1965), Peto & Peto(1972), Tarone & Ware(1977), Prentice(1978), Harrington & Fleming(1982), Fleming & Harrington(1991)]

Our objective is to propose methodology for the use of Fleming and Harrington test in the setting of clinical trials.

The constraints of a Clinical trial design are

- The test has to be chosen before the knowledge of the data.
- All the parameters have to be chosen by the investigator.
- The sample size can (and has to) be calculated.
Many tests have been defined to detect late (early) effects. [Gehan(1965), Peto & Peto(1972), Tarone & Ware(1977), Prentice(1978), Harrington & Fleming(1982), Fleming & Harrington(1991)]

Our objective is to propose methodology for the use of Fleming and Harrington test in the setting of clinical trials.

The constraints of a Clinical trial design are

1. The test has to be chosen before the knowledge of the data. All the parameters have to be chosen by the investigator.
2. The sample size can (and has to) be calculated.
Many tests have been defined to detect late (early) effects.


Our objective is to propose methodology for the use of Fleming and Harrington test in the setting of clinical trials.

The constraints of a Clinical trial design are

1. The test has to be chosen before the knowledge of the data. All the parameters have to be chosen by the investigator.

2. The sample size can (and has to) be calculated.
Many tests have been defined to detect late (early) effects. [Gehan(1965), Peto & Peto(1972), Tarone & Ware(1977), Prentice(1978), Harrington & Fleming(1982), Fleming & Harrington(1991)]

Our objective is to propose **methodology for the use** of Fleming and Harrington test in the setting of clinical trials.

The constraints of a Clinical trial design are

1. The test has to be chosen before the knowledge of the data. All the parameters have to be chosen by the investigator.

2. The sample size can (and has to) be calculated.
Let us recall the notations of survival data analysis:

- \( T \): a non-negative random variable "duration between the origin date and the time of occurrence of some specific event"
- \( F \): the cumulative distribution function associated to \( T \)
- \( S = 1 - F \): the survival function associated to \( T \)
- \( \lambda \): the hazard function associated to \( T \)

\[
\lambda(t) = \lim_{dt \to 0} \frac{\mathbb{P}[t \leq T < t + dt | T \geq t]}{dt}.
\]

\( \Lambda(t) = \int_0^t \lambda(s)ds \): the cumulative hazard function associated to \( T \)

We consider **right-censoring**: \( C \) independent of \( T \)

For each subject \( i = 1, \ldots, n \), we observe:

\[
X_i = T_i \wedge C_i,
\]

and

\[
\delta_i = \mathbb{I}\{T_i \leq C_i\}.
\]
The processes associated with survival data analysis are:

- **Number of failures at** $t$ :

  \[ N_n(t) = \sum_{i=1}^{n} \mathbb{I}\{X_i \leq t, \delta_i = 1\} \]

- **Number of subjects at risk at** $t^-$ :

  \[ Y_n(t) = \sum_{i=1}^{n} \mathbb{I}\{X_i \geq t\} \]
A test is said to be of class $\mathcal{K}$ if its discriminant function can be written

$$LR_{W_n}(t) = \int_0^t W_n(s) \sqrt{\frac{n_P + n_T}{n_P n_T}} Y_{n_p}^P(s) Y_{n_T}^T(s) \left[ \frac{dN_{n_p}^P(s)}{Y_{n_p}^P(s)} - \frac{dN_{n_T}^T(s)}{Y_{n_T}^T(s)} \right],$$

with $W_n$ an adapted bounded non-negative predictable process.

- $W_n(s) = 1$ : Logrank test
- $W_n(s) = Y_n(s)$ : Gehan's test
- $W_n(s) = \hat{S}(s)^p, p \geq 0$ : Fleming and Harrington test for early effect
- $W_n(s) = (1 - \hat{S}(s))^q, q \geq 0$ : Fleming and Harrington test for late effect

$N_n(s)$ is the number of failures at $s$

$Y_n(s)$ is the number of subjects at risk at $s$

$\hat{S}(s)$ is the K-M estimator of survival at time $s$ for mixed samples.
A test is said to be of class $\mathcal{K}$ if its discriminant function can be written

$$LR_{W_n}(t) = \int_0^t W_n(s) \sqrt{\frac{n_P + n_T}{n_P n_T}} \frac{Y_{n_P}^P(s) Y_{n_T}^T(s)}{Y_n(s)} \left[ \frac{dN_{n_P}(s)}{Y_{n_P}^P(s)} - \frac{dN_{n_T}(s)}{Y_{n_T}^T(s)} \right],$$

with $W_n$ an adapted bounded non-negative predictable process.

- $W_n(s) = 1$ : Logrank test
- $W_n(s) = Y_n(s)$ : Gehan’s test
- $W_n(s) = \hat{S}(s)^p$, $p \geq 0$ : Fleming and Harrington test for early effect
- $W_n(s) = (1 - \hat{S}(s))^q$, $q \geq 0$ : Fleming and Harrington test for late effect

$N_n(s)$ is the number of failures at $s$
$Y_n(s)$ is the number of subjects at risk at $s$
$\hat{S}(s)$ is the K-M estimator of survival at time $s$ for mixed samples.
Definition

A test is said to be of class $\mathcal{K}$ if its discriminant function can be written

$$LR_{W_n}(t) = \int_0^t W_n(s) \sqrt{\frac{n_P + n_T}{n_P n_T}} \frac{Y_{n_P}^P(s) Y_{n_T}^T(s)}{Y_n(s)} \left[ \frac{dN_{n_P}^P(s)}{Y_{n_P}^P(s)} - \frac{dN_{n_T}^T(s)}{Y_{n_T}^T(s)} \right],$$

with $W_n$ an adapted bounded non-negative predictable process.

- $W_n(s) = 1$ : Logrank test
- $W_n(s) = Y_n(s)$ : Gehan’s test
- $W_n(s) = \hat{S}(s)^p$, $p \geq 0$ : Fleming and Harrington test for early effect
- $W_n(s) = (1 - \hat{S}(s))^q$, $q \geq 0$ : Fleming and Harrington test for late effect

$N_n(s)$ is the number of failures at $s$

$Y_n(s)$ is the number of subjects at risk at $s$

$\hat{S}(s)$ is the K-M estimator of survival at time $s$ for mixed samples.
A test is said to be of class $\mathcal{K}$ if its discriminant function can be written

$$LR_{W_n}(t) = \int_0^t W_n(s) \sqrt{\frac{n_P + n_T}{n_P n_T}} \frac{Y_{n_P}^P(s) Y_{n_T}^T(s)}{Y_n(s)} \left[ \frac{dN_{n_P}^P(s)}{Y_{n_P}^P(s)} - \frac{dN_{n_T}^T(s)}{Y_{n_T}^T(s)} \right],$$

with $W_n$ an adapted bounded non-negative predictable process.

- $W_n(s) = 1$ : Logrank test
- $W_n(s) = Y_n(s)$ : Gehan’s test
- $W_n(s) = \hat{S}(s)^p, \ p \geq 0$ : Fleming and Harrington test for early effect
- $W_n(s) = (1 - \hat{S}(s))^q, \ q \geq 0$ : Fleming and Harrington test for late effect

$N_n(s)$ is the number of failures at $s$
$Y_n(s)$ is the number of subjects at risk at $s^-$
$\hat{S}(s)$ is the K-M estimator of survival at time $s$ for mixed samples.
Definition

A test is said to be of class $\mathcal{K}$ if its discriminant function can be written

$$LR_{W_n}(t) = \int_0^t W_n(s) \sqrt{\frac{n_p + n_T}{n_p n_T}} \frac{Y_{n_p}^P(s) Y_{n_T}^T(s)}{Y_n(s)} \left[ \frac{dN_{n_p}^P(s)}{Y_{n_p}^P(s)} - \frac{dN_{n_T}^T(s)}{Y_{n_T}^T(s)} \right],$$

with $W_n$ an adapted bounded non-negative predictable process.

- $W_n(s) = 1$ : Logrank test
- $W_n(s) = Y_n(s)$ : Gehan’s test
- $W_n(s) = \hat{S}(s)^p, p \geq 0$ : Fleming and Harrington test for early effect
- $W_n(s) = (1 - \hat{S}(s))^q, q \geq 0$ : Fleming and Harrington test for late effect

$N_n(s)$ is the number of failures at $s$
$Y_n(s)$ is the number of subjects at risk at $s$
$\hat{S}(s)$ is the K-M estimator of survival at time $s$ for mixed samples.
**Definition**

A test is said to be of class $\mathcal{K}$ if its discriminant function can be written

$$LR_{W_n}(t) = \int_0^t W_n(s) \sqrt{\frac{n_P + n_T}{n_P n_T}} \frac{Y_{n_P}^P(s) Y_{n_T}^T(s)}{Y_n(s)} \left[ \frac{dN_{n_P}^P(s)}{Y_{n_P}^P(s)} - \frac{dN_{n_T}^T(s)}{Y_{n_T}^T(s)} \right],$$

with $W_n$ an adapted bounded non-negative predictable process.

- $W_n(s) = 1$ : Logrank test
- $W_n(s) = Y_n(s)$ : Gehan’s test
- $W_n(s) = \hat{S}(s)^p$, $p \geq 0$ : Fleming and Harrington test for early effect
- $W_n(s) = (1 - \hat{S}(s))^q$, $q \geq 0$ : Fleming and Harrington test for late effect

$N_n(s)$ is the number of failures at $s$
$Y_n(s)$ is the number of subjects at risk at $s$
$\hat{S}(s)$ is the K-M estimator of survival at time $s$ for mixed samples.
A test is said to be of class $\mathcal{K}$ if its discriminant function can be written

$$LR_{W_n}(t) = \int_0^t W_n(s) \sqrt{\frac{n_P + n_T}{n_P n_T}} \frac{Y_{n_P}^P(s) Y_{n_T}^T(s)}{Y_n(s)} \left[ \frac{dN_{n_P}^P(s)}{Y_{n_P}^P(s)} - \frac{dN_{n_T}^T(s)}{Y_{n_T}^T(s)} \right],$$

with $W_n$ an adapted bounded non-negative predictable process.

- $W_n(s) = 1$ : Logrank test
- $W_n(s) = Y_n(s)$ : Gehan’s test
- $W_n(s) = \hat{S}(s)^p$, $p \geq 0$ : Fleming and Harrington test for early effect
- $W_n(s) = (1 - \hat{S}(s))^q$, $q \geq 0$ : Fleming and Harrington test for late effect

$N_n(s)$ is the number of failures at $s$
$Y_n(s)$ is the number of subjects at risk at $s$
$\hat{S}(s)$ is the K-M estimator of survival at time $s$ for mixed samples.
Now, we focus on the Fleming and Harrington test for late effects:

\[
LR^q(t) = \int_0^t (1 - \hat{S}(s))^q \left( \frac{n_P + n_T}{n_P n_T} \right)^{1/2} \frac{Y_{n_P}^P(s) Y_{n_T}^T(s)}{Y_n(s)} \times \Delta_n(s),
\]

where \( \Delta_n(s) \) denotes the distance between hazard functions at time \( s \)

\[
\Delta_n(s) = \frac{dN_{n_P}^P(s)}{Y_{n_P}^P(s)} - \frac{dN_{n_T}^T(s)}{Y_{n_T}^T(s)}.
\]

Why is it adapted to the detection of late effects?

For \( q \) fixed, \( t \rightarrow (1 - \hat{S}(t))^q \) is increasing.

\[\implies\] the more time goes on, the more weight is given to distances.
Now, we focus on the Fleming and Harrington test for late effects:

$$LR^q(t) = \int_0^t (1 - \hat{S}(s))^q \left( \frac{n_P + n_T}{n_P n_T} \right)^{1/2} \frac{Y_{n_P}(s) Y_{n_T}(s)}{Y_n(s)} \times \Delta_n(s),$$

where $\Delta_n(s)$ denotes the distance between hazard functions at time $s$

$$\Delta_n(s) = \frac{dN_{n_P}(s)}{Y_{n_P}(s)} - \frac{dN_{n_T}(s)}{Y_{n_T}(s)}.$$

Why is it adapted to the detection of late effects?

For $q$ fixed, $t \to (1 - \hat{S}(t))^q$ is increasing.

$\implies$ the more time goes on, the more weight is given to distances.
Now, we focus on the Fleming and Harrington test for late effects:

\[ LR^q(t) = \int_0^t (1 - \hat{S}(s))^q \left( \frac{n_P + n_T}{n_P n_T} \right)^{1/2} \frac{Y_{n_P}^P(s) Y_{n_T}^T(s)}{Y_n(s)} \times \Delta_n(s), \]

where \( \Delta_n(s) \) denotes the distance between hazard functions at time \( s \)

\[ \Delta_n(s) = \frac{dN_{n_P}^P(s)}{Y_{n_P}^P(s)} - \frac{dN_{n_T}^T(s)}{Y_{n_T}^T(s)}. \]

Why is it adapted to the detection of late effects?

For \( q \) fixed, \( t \to (1 - \hat{S}(t))^q \) is increasing.

\( \rightarrow \) the more time goes on, the more weight is given to distances.
Now, we focus on the Fleming and Harrington test for late effects:

\[ LR^q(t) = \int_0^t (1 - \hat{S}(s))^q \left( \frac{n_P + n_T}{n_P n_T} \right)^{1/2} \frac{Y_{nP}^P(s) Y_{nT}^T(s)}{Y_n(s)} \times \Delta_n(s), \]

where \( \Delta_n(s) \) denotes the distance between hazard functions at time \( s \)

\[ \Delta_n(s) = \frac{dN_{nP}^P(s)}{Y_{nP}^P(s)} - \frac{dN_{nT}^T(s)}{Y_{nT}^T(s)}. \]

Why is it adapted to the detection of late effects?

For \( q \) fixed, \( t \rightarrow (1 - \hat{S}(t))^q \) is increasing.

\( \Rightarrow \) the more time goes on, the more weight is given to distances.
Fleming and Harrington test

**But** for $t$ fixed, $q \rightarrow (1 - \hat{S}(t))^q$ is decreasing.

$\implies$ Bad control of test

$\implies$ Hard to define the value of $q$

$\implies$ The value of $q$ must not be too large.
But for $t$ fixed, $q \rightarrow (1 - \hat{S}(t))^q$ is decreasing.

$\Rightarrow$ Bad control of test

$\Rightarrow$ Hard to define the value of $q$

$\Rightarrow$ The value of $q$ must not be too large.
But for $t$ fixed, $q \to (1 - \hat{S}(t))^q$ is decreasing.

$\implies$ Bad control of test

$\implies$ Hard to define the value of $q$

$\implies$ The value of $q$ must not be too large.
But for $t$ fixed, $q \rightarrow (1 - \hat{S}(t))^q$ is decreasing.

$\implies$ Bad control of test

$\implies$ Hard to define the value of $q$

$\implies$ The value of $q$ must not be too large.
**But** for $t$ fixed, $q \to (1 - \hat{S}(t))^q$ is decreasing.

$\implies$ Bad control of test

$\implies$ Hard to define the value of $q$

$\implies$ The value of $q$ must not be too large.
**But** for \( t \) fixed, \( q \rightarrow (1 - \hat{S}(t))^q \) is decreasing.

\[ \Rightarrow \text{Bad control of test} \]

\[ \Rightarrow \text{Hard to define the value of } q \]

\[ \Rightarrow \text{The value of } q \text{ must not be too large.} \]
But for $t$ fixed, $q \to (1 - \hat{S}(t))^q$ is decreasing.

$\implies$ Bad control of test

$\implies$ Hard to define the value of $q$

$\implies$ The value of $q$ must not be too large.
**Fleming and Harrington test**

But for $t$ fixed, $q \to (1 - \hat{S}(t))^q$ is decreasing.

$\Rightarrow$ Bad control of test

$\Rightarrow$ Hard to define the value of $q$

$\Rightarrow$ The value of $q$ must not be too large.
**Fleming and Harrington test**

**But** for $t$ fixed, $q \rightarrow (1 - \hat{S}(t))^q$ is decreasing.

$\Rightarrow$ Bad control of test

$\Rightarrow$ Hard to define the value of $q$

$\Rightarrow$ The value of $q$ must not be too large.
**Fleming and Harrington test**

**But** for $t$ fixed, $q \rightarrow (1 - \hat{S}(t))^q$ is decreasing.

$\Rightarrow$ Bad control of test

$\Rightarrow$ Hard to define the value of $q$

$\Rightarrow$ The value of $q$ must not be too large.

\[
\begin{array}{c|c|c}
 & \text{Treat 1} & \text{Treat 2} \\
\hline
\text{Logrank} & \mathcal{H}_0 & \mathcal{H}_0 \\
\text{FH(1)} & \mathcal{H}_1 & \mathcal{H}_0 \\
\text{FH(4)} & \mathcal{H}_0 & \mathcal{H}_1 \\
\end{array}
\]
In our setting, the assumptions are

\[ \begin{align*}
\mathcal{H}_0 &: F^T = F^P = F_{\theta_0}, \\
\mathcal{H}_1 &: F^T = F_{\theta_T} \quad \text{et} \quad F^P = F_{\theta_P}
\end{align*} \]

for \( i = T, P \). \( (\theta^i_{n_i}) \) is a sequence of \( \Theta \subset \mathbb{R} \).
Theorem (Quality of the test - asymptotic distribution)

Let $LR_{W_n}$, a statistic in class $\mathcal{K}$ such that

$$W_n(s) \xrightarrow{a.s.} n \to \infty w(s)$$

Then, under $\mathcal{H}_1$,

$$LR_{W_n} - \sqrt{n} \mu_{G} \xrightarrow{n \to \infty} \mathcal{G},$$

where

$$\mu_G : t \to \int_0^t k(s) \sqrt{a^P a^T (d\Lambda_{\theta P}(s) - d\Lambda_{\theta T})}(s),$$

where

$$k(s) = w(s) \frac{\pi^P(s) \pi^T(s)}{a^P \pi^P(s) + a^T \pi^T(s)}$$

and $\mathcal{G}$ is a centred Gaussian process with covariance function :

$$(t_1, t_2) \to \int_0^{t_1 \wedge t_2} k^2(s) \left[ \frac{a^T}{\pi^P(s)} (1 - \Delta\Lambda_{\theta P}(s)) d\Lambda_{\theta P}(s) + \frac{a^P}{\pi^T(s)} (1 - \Delta\Lambda_{\theta T}(s)) d\Lambda_{\theta T}(s) \right].$$
Pitman’s Asymptotic Relative Efficiency

How to compare two tests?
Consider two sequences of statistics \((T_n)\) and \((V_n)\) based on \(n\) observations testing the assumptions

\[
\begin{align*}
\mathcal{H}_0 & : \theta = \theta_0, \\
\mathcal{H}_1 & : \theta \neq \theta_0.
\end{align*}
\]

Denote \(N_T(\alpha, \beta, \theta)\) the sample size necessary for \(T\) to attain the power \(\beta\) under the level \(\alpha\) and the alternative value of parameter \(\theta\).

\[
RE(T, V) = \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)}
\]

● permits to compare two tests and is universally acknowledged.
● But explicit computation of \(N_T(\alpha, \beta, \theta)\) extremely difficult.

A solution is to consider Asymptotic Relative Efficiency (ARE)
How to compare two tests?
Consider two sequences of statistics \( (T_n) \) and \( (V_n) \) based on \( n \) observations testing the assumptions
\[
\begin{align*}
\mathcal{H}_0 & : \theta = \theta_0, \\
\mathcal{H}_1 & : \theta \neq \theta_0.
\end{align*}
\]
Denote \( N_T(\alpha, \beta, \theta) \) the sample size necessary for \( T \) to attain the power \( \beta \) under the level \( \alpha \) and the alternative value of parameter \( \theta \).

\[
RE(T, V) = \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)}
\]

- permits to compare two tests and is universally acknowledged.
- But explicit computation of \( N_T(\alpha, \beta, \theta) \) extremely difficult.
A solution is to consider Asymptotic Relative Efficiency (ARE)
How to compare two tests?
Consider two sequences of statistics \((T_n)\) and \((V_n)\) based on \(n\) observations testing the assumptions

\[
\begin{align*}
\mathcal{H}_0 & : \theta = \theta_0, \\
\mathcal{H}_1 & : \theta \neq \theta_0.
\end{align*}
\]

Denote \(N_T(\alpha, \beta, \theta)\) the sample size necessary for \(T\) to attain the power \(\beta\) under the level \(\alpha\) and the alternative value of parameter \(\theta\).

\[
RE(T, V) = \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)}
\]

permits to compare two tests and is universally acknowledged.

But explicit computation of \(N_T(\alpha, \beta, \theta)\) extremely difficult.

A solution is to consider Asymptotic Relative Efficiency (ARE)
How to compare two tests?

Consider two sequences of statistics \((T_n)\) and \((V_n)\) based on \(n\) observations testing the assumptions

\[
\begin{align*}
\mathcal{H}_0 & : \theta = \theta_0, \\
\mathcal{H}_1 & : \theta \neq \theta_0.
\end{align*}
\]

Denote \(N_T(\alpha, \beta, \theta)\) the sample size necessary for \(T\) to attain the power \(\beta\) under the level \(\alpha\) and the alternative value of parameter \(\theta\).

\[
RE(T, V) = \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)}
\]

- permits to compare two tests and is universally acknowledged.
- But explicit computation of \(N_T(\alpha, \beta, \theta)\) extremely difficult.

A solution is to consider Asymptotic Relative Efficiency (ARE)
How to compare two tests?
Consider two sequences of statistics \((T_n)\) and \((V_n)\) based on \(n\) observations testing the assumptions

\[
\begin{align*}
\mathcal{H}_0 & : \theta = \theta_0, \\
\mathcal{H}_1 & : \theta \neq \theta_0.
\end{align*}
\]

Denote \(N_T(\alpha, \beta, \theta)\) the sample size necessary for \(T\) to attain the power \(\beta\) under the level \(\alpha\) and the alternative value of parameter \(\theta\).

\[
RE(T, V) = \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)}
\]

- permits to compare two tests and is universally acknowledged.
- But explicit computation of \(N_T(\alpha, \beta, \theta)\) extremely difficult.
A solution is to consider Asymptotic Relative Efficiency (ARE)
How to compare two tests?
Consider two sequences of statistics \((T_n)\) and \((V_n)\) based on \(n\) observations testing the assumptions
\[
\begin{align*}
\mathcal{H}_0 & : \theta = \theta_0, \\
\mathcal{H}_1 & : \theta \neq \theta_0.
\end{align*}
\]
Denote \(N_T(\alpha, \beta, \theta)\) the sample size necessary for \(T\) to attain the power \(\beta\) under the level \(\alpha\) and the alternative value of parameter \(\theta\).

\[
RE(T, V) = \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)}
\]

- permits to compare two tests and is universally acknowledged.
- But explicit computation of \(N_T(\alpha, \beta, \theta)\) extremely difficult.
A solution is to consider Asymptotic Relative Efficiency (ARE)
Pitman’s Asymptotic Relative Efficiency

How to compare two tests?
Consider two sequences of statistics \((T_n)\) and \((V_n)\) based on \(n\) observations testing the assumptions

\[
\begin{align*}
\mathcal{H}_0 & : \theta = \theta_0, \\
\mathcal{H}_1 & : \theta \neq \theta_0.
\end{align*}
\]

Denote \(N_T(\alpha, \beta, \theta)\) the sample size necessary for \(T\) to attain the power \(\beta\) under the level \(\alpha\) and the alternative value of parameter \(\theta\).

\[
RE(T, V) = \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)}
\]

- permits to compare two tests and is universally acknowledged.
- But explicit computation of \(N_T(\alpha, \beta, \theta)\) extremely difficult.

A solution is to consider Asymptotic Relative Efficiency (ARE)

\[
\begin{align*}
\lim_{\alpha \to 0} \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)} & \text{ Bahadur} \\
\lim_{\beta \to 1} \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)} & \text{ Hodges-Lehmann} \\
\lim_{\theta \to \theta_0} \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)} & \text{ Pitman}
\end{align*}
\]
How to compare two tests?
Consider two sequences of statistics \((T_n)\) and \((V_n)\) based on \(n\) observations testing the assumptions

\[
\begin{align*}
H_0 & : \theta = \theta_0, \\
H_1 & : \theta \neq \theta_0.
\end{align*}
\]

Denote \(N_T(\alpha, \beta, \theta)\) the sample size necessary for \(T\) to attain the power \(\beta\) under the level \(\alpha\) and the alternative value of parameter \(\theta\).

\[
RE(T, V) = \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)}
\]

- permits to compare two tests and is universally acknowledged.
- But explicit computation of \(N_T(\alpha, \beta, \theta)\) extremely difficult.

A solution is to consider Asymptotic Relative Efficiency (ARE)

\[
\begin{align*}
\lim_{\alpha \to 0} \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)} & \text{ Bahadur} \\
\lim_{\beta \to 1} \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)} & \text{ Hodges-Lehmann} \\
\lim_{\theta \to \theta_0} \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)} & \text{ Pitman}
\end{align*}
\]
Why Pitman’s ARE?

**Theorem ([van der Vaart(1998)])**

Consider the sequence of assumptions

\[
\begin{cases}
\mathcal{H}_0 & : \theta = \theta_0, \\
\mathcal{H}_1 & : \theta = \theta_\nu = \theta_0 + \frac{c}{\sqrt{\nu}}.
\end{cases}
\]

Assume that for any \(\theta\),

\[
\sqrt{n} \frac{T_n - \mu_T(\theta)}{\sigma_T(\theta)} \xrightarrow{n \to \infty} \mathcal{N}(0, 1) \quad \text{and} \quad \sqrt{n} \frac{V_n - \mu_V(\theta)}{\sigma_V(\theta)} \xrightarrow{n \to \infty} \mathcal{N}(0, 1).
\]

Under suitable assumptions,

\[
\text{ARE}(T, V) = \left( \frac{\mu_V(\theta_0)/\sigma_V(\theta_0)}{\mu_T(\theta_0)/\sigma_T(\theta_0)} \right)^2.
\]
Why Pitman’s ARE?

**Theorem ([van der Vaart(1998)])**

*Consider the sequence of assumptions*

\[
\begin{align*}
\mathcal{H}_0 & : \theta = \theta_0, \\
\mathcal{H}_1 & : \theta = \theta_\nu = \theta_0 + \frac{c}{\sqrt{\nu}}.
\end{align*}
\]

Assume that for any \(\theta\),

\[
\sqrt{n} \frac{T_n - \mu_T(\theta)}{\sigma_T(\theta)} \xrightarrow{n \to \infty} \mathcal{N}(0,1) \quad \text{and} \quad \sqrt{n} \frac{V_n - \mu_V(\theta)}{\sigma_V(\theta)} \xrightarrow{n \to \infty} \mathcal{N}(0,1).
\]

Under suitable assumptions,

\[
\text{ARE}(T, V) = \left( \frac{\mu_V(\theta_0)/\sigma_V(\theta_0)}{\mu_T(\theta_0)/\sigma_T(\theta_0)} \right)^2.
\]
Why Pitman’s ARE?

Theorem ([van der Vaart(1998)])

Consider the sequence of assumptions

\[
\begin{align*}
\mathcal{H}_0 & : \theta = \theta_0, \\
\mathcal{H}_1 & : \theta = \theta_\nu = \theta_0 + \frac{\xi}{\sqrt{\nu}}.
\end{align*}
\]

Assume that for any \( \theta \),

\[
\sqrt{n} \frac{T_n - \mu_T(\theta)}{\sigma_T(\theta)} \xrightarrow{n \to \infty} \mathcal{N}(0, 1) \quad \text{and} \quad \sqrt{n} \frac{V_n - \mu_V(\theta)}{\sigma_V(\theta)} \xrightarrow{n \to \infty} \mathcal{N}(0, 1).
\]

Under suitable assumptions,

\[
\text{ARE}(T, V) = \left( \frac{\mu_V(\theta_0)/\sigma_V(\theta_0)}{\mu_T(\theta_0)/\sigma_T(\theta_0)} \right)^2.
\]
Why Pitman’s ARE?

Theorem ([van der Vaart(1998)])

Consider the sequence of assumptions

\[\begin{align*}
\mathcal{H}_0 & : \theta = \theta_0, \\
\mathcal{H}_1 & : \theta = \theta_0 + \frac{c}{\sqrt{\nu}}.
\end{align*}\]

Assume that for any \(\theta\),

\[\sqrt{n} \frac{T_n - \mu_T(\theta)}{\sigma_T(\theta)} \xrightarrow{n \to \infty} \mathcal{N}(0, 1) \quad \text{and} \quad \sqrt{n} \frac{V_n - \mu_V(\theta)}{\sigma_V(\theta)} \xrightarrow{n \to \infty} \mathcal{N}(0, 1).\]

Under suitable assumptions,

\[\text{ARE}(T, V) = \left(\frac{\mu_V(\theta_0)/\sigma_V(\theta_0)}{\mu_T(\theta_0)/\sigma_T(\theta_0)}\right)^2.\]
ARE of tests of class and Shift Assumptions

In our setting, the assumptions are

\[ \begin{align*}
\mathcal{H}_0 & : F^T = F^P = F_{\theta_0}, \\
\mathcal{H}_1 & : F^T = F_{\theta_n^T} \quad \text{et} \quad F^P = F_{\theta_n^P},
\end{align*} \]

with

\[ \begin{align*}
\theta_n^P & = \theta_0 + c \left( \frac{n_T}{n_P(n_P + n_T)} \right)^{1/2}, \\
\theta_n^T & = \theta_0 - c \left( \frac{n_P}{n_T(n_T + n_T)} \right)^{1/2}.
\end{align*} \]

Let us restrain ourselves to shift assumption of the form

\[ F_{\theta}(t) = \Psi(g(t) + \theta), \quad \theta \in \Theta, \]

with

- \( g \) is a differentiable non-decreasing function from \([0, \infty[ \) to \( -\infty, u^+ \]
  with \( u^+ \in \bar{\mathbb{R}} \).
- \( \Psi \) is a continuous cumulative distribution function having a continuous second derivative.
ARE of tests of class and Shift Assumptions

In our setting, the assumptions are

\[
\begin{align*}
\mathcal{H}_0 & : F^T = F^P = F_{\theta_0}, \\
\mathcal{H}_1 & : F^T = F_{\theta^T_{nT}} \text{ et } F^P = F_{\theta^P_{nP}},
\end{align*}
\]

with

\[
\begin{align*}
\theta^P_{nP} &= \theta_0 + c \left( \frac{n_T}{n_P(n_P + n_T)} \right)^{1/2}, \\
\theta^T_{nT} &= \theta_0 - c \left( \frac{n_P}{n_T(n_P + n_T)} \right)^{1/2}.
\end{align*}
\]

Let us restrain ourselves to shift assumption of the form

\[
F_\theta(t) = \Psi(g(t) + \theta), \quad \theta \in \Theta,
\]

with

- \(g\) is a differentiable non-decreasing function from \([0, \infty[\) to \(]-\infty, u^+ [\)
  with \(u^+ \in \bar{\mathbb{R}}\).

- \(\Psi\) is a continuous cumulative distribution function having a continuous second derivative.
ARE of tests of class and Shift Assumptions

To maximise the ARE, it is enough to maximise \( AE = \frac{\mu}{\sigma} \). An application of Cauchy-Schwarz Theorem yields to:

**Theorem ([Gill(1980)])**

The limit weights of the statistic in the class \( K \) for which the asymptotic efficiency is maximal to test the shift assumptions are proportional together and verify, for all \( t \in \mathbb{R}^+ \):

\[
    w(t) = P'[\psi] \circ \psi^{-1} \circ F_{\theta_0}(t) \quad \text{where} \quad P(\Phi) = \ln \left( \frac{\Phi'}{1 - \Phi} \right). \tag{1}
\]

\[ \text{Weight} \underset{(1)}{\iff} \text{Pattern of the optimal assumptions} \]
To maximise the ARE, it is enough to maximise $AE = \frac{\mu}{\sigma}$. An application of Cauchy-Schwarz Theorem yields to:

**Theorem ([Gill(1980)])**

The limit weights of the statistic in the class $\mathcal{K}$ for which the asymptotic efficiency is maximal to test the shift assumptions are proportional together and verify, for all $t \in \mathbb{R}^+$:

$$w(t) = P'[\psi] \circ \psi^{-1} \circ F_{\theta_0}(t) \quad \text{where} \quad P(\Phi) = \ln \left( \frac{\Phi'}{1 - \Phi} \right). \quad (1)$$

**Weight $(1)$ Pattern of the optimal assumptions**
To maximise the ARE, it is enough to maximise $AE = \frac{\mu}{\sigma}$. An application of Cauchy-Schwarz Theorem yields to:

**Theorem ([Gill(1980)])**

The limit weights of the statistic in the class $\mathcal{K}$ for which the asymptotic efficiency is maximal to test the shift assumptions are proportional together and verify, for all $t \in \mathbb{R}^+$:

$$w(t) = P'[\psi] \circ \psi^{-1} \circ F_{\theta_0}(t) \quad \text{where} \quad P(\Phi) = \ln \left( \frac{\Phi'}{1 - \Phi} \right).$$

(1)

Weight $\overset{(1)}{\Longleftrightarrow}$ Pattern of the optimal assumptions
Pattern of the optimal assumptions

Denote $\Delta = \theta^P - \theta^T$.

- For Logrank test $W = 1$ corresponds to the pattern
  \[
  \begin{align*}
  \mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
  \mathcal{H}_1 & : \lambda_T = \lambda_P e^\Delta.
  \end{align*}
  \]

- For Fleming and Harrington for early effect $W = \hat{S}^p$ corresponds to the pattern
  \[
  \begin{align*}
  \mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
  \mathcal{H}_1 & : \lambda_T = \lambda_P e^\Delta \left[ (S_P)^p + [1 - (S_P)^p] e^\Delta \right]^{-1}.
  \end{align*}
  \]

- For Fleming and Harrington for late effect $W = (1 - \hat{S})^q$ corresponds to the pattern
  \[
  \begin{align*}
  \mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
  \mathcal{H}_1 & : \lambda_T = \lambda_P \varphi^q(S_P, \Delta).
  \end{align*}
  \]
Denote $\Delta = \theta^P - \theta^T$.

- For **Logrank test** $W = 1$ corresponds to the pattern

$$
\begin{align*}
\mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
\mathcal{H}_1 & : \lambda_T = \lambda_P e^{\Delta}.
\end{align*}
$$

- For **Fleming and Harrington** for early effect $W = \hat{S}^p$ corresponds to the pattern

$$
\begin{align*}
\mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
\mathcal{H}_1 & : \lambda_T = \lambda_P e^{\Delta} [(S_P)^p + [1 - (S_P)^p] e^{\Delta}]^{-1}.
\end{align*}
$$

- For **Fleming and Harrington** for late effect $W = (1 - \hat{S})^q$ corresponds to the pattern

$$
\begin{align*}
\mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
\mathcal{H}_1 & : \lambda_T = \lambda_P \phi^q(S_P, \Delta).
\end{align*}
$$
Pattern of the optimal assumptions

Denote $\Delta = \theta^P - \theta^T$.

- For **Logrank test** $W = 1$ corresponds to the pattern
  
  $$
  \begin{align*}
  \mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
  \mathcal{H}_1 & : \lambda_T = \lambda_P e^\Delta.
  \end{align*}
  $$

- For **Fleming and Harrington for early effect** $W = \hat{S}^p$ corresponds to the pattern
  
  $$
  \begin{align*}
  \mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
  \mathcal{H}_1 & : \lambda_T = \lambda_P e^\Delta [(S_P)^p + [1 - (S_P)^p] e^\Delta]^{-1}.
  \end{align*}
  $$

- For **Fleming and Harrington for late effect** $W = (1 - \hat{S})^q$ corresponds to the pattern
  
  $$
  \begin{align*}
  \mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
  \mathcal{H}_1 & : \lambda_T = \lambda_P \phi^q(S_P, \Delta).
  \end{align*}
  $$
Pattern of the optimal assumptions

Denote $\Delta = \theta^P - \theta^T$.

- For **Logrank test** $W = 1$ corresponds to the pattern
  \[
  \begin{cases}
  \mathcal{H}_0 : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
  \mathcal{H}_1 : \lambda_T = \lambda_P e^\Delta.
  \end{cases}
  \]

- For **Fleming and Harrington for early effect** $W = \hat{S}^p$ corresponds to the pattern
  \[
  \begin{cases}
  \mathcal{H}_0 : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
  \mathcal{H}_1 : \lambda_T = \lambda_P e^\Delta [(S_P)^p + [1 - (S_P)^p]e^\Delta]^{-1}.
  \end{cases}
  \]

- For **Fleming and Harrington for late effect** $W = (1 - \hat{S})^q$ corresponds to the pattern
  \[
  \begin{cases}
  \mathcal{H}_0 : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
  \mathcal{H}_1 : \lambda_T = \lambda_P \phi^q(S_P, \Delta).
  \end{cases}
  \]
Pattern of the optimal assumptions
Proportional and early effects
Pattern of the optimal assumptions

Late effects

**FIGURE:** Hazard functions (left hand side) and Survival functions (right hand side)
Fleming and Harrington test is optimal to test

$$\begin{align*}
\mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
\mathcal{H}_1 & : \lambda_T = \lambda_P \phi^q(S_P, \Delta).
\end{align*}$$

(2)

The knowledge of $\phi^q$ allows us to run simulation studies.

- Consider $n = 2000$, $\tau = 5$ years, $S_P(5) = 0.80$, $r = \frac{S^T(5) - S^P(5)}{1 - S^P(5)} = 20\%$.
- Consider an exponential model in the Placebo arm.
Fleming and Harrington test is optimal to test

\[
\begin{align*}
\mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
\mathcal{H}_1 & : \lambda_T = \lambda_P \phi^q(S_P, \Delta).
\end{align*}
\]  

(2)

The knowledge of \(\phi^q\) allows us to run simulation studies.

- Consider \(n = 2000\), \(\tau = 5\) years, \(S_P(5) = 0.80\), \(r = \frac{S_T(5) - S_P(5)}{1 - S_P(5)} = 20\%\).

- Consider an exponential model in the Placebo arm.
Fleming and Harrington test is optimal to test

\[
\begin{align*}
\mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
\mathcal{H}_1 & : \lambda_T = \lambda_P \phi^q(S_P, \Delta).
\end{align*}
\]

The knowledge of $\phi^q$ allows us to run simulation studies.

- Consider $n = 2000$, $\tau = 5$ years, $S_P(5) = 0.80$, $r = \frac{S^T(5) - S_P(5)}{1 - S_P(5)} = 20\%$.

- Consider an exponential model in the Placebo arm.
Fleming and Harrington test is optimal to test

\[
\begin{align*}
\mathcal{H}_0 & : \lambda_T = \lambda_P = \lambda_{\theta_0}, \\
\mathcal{H}_1 & : \lambda_T = \lambda_P \phi^q(S_P, \Delta).
\end{align*}
\]

(2)

The knowledge of \(\phi^q\) allows us to run simulation studies.

- Consider \(n = 2000\), \(\tau = 5\) years, \(S_P(5) = 0.80\), \(r = \frac{S_T(5) - S_P(5)}{1 - S_P(5)} = 20\%\).

- Consider an exponential model in the Placebo arm.
Consider data generated under $q = 4$

<table>
<thead>
<tr>
<th></th>
<th>Logrank</th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.59</td>
<td>0.88</td>
<td>0.94</td>
<td>0.95</td>
<td>0.96</td>
<td>0.94</td>
</tr>
</tbody>
</table>

- The power of the F-H test is very good.
- The power of the logrank test is very low under non-proportional hazards assumption.
- The test is not sensitive to a variation of the value $q$.
  ⇒ Very reassuring for its use in clinical trials

Consider data generated under $q = 0$ (proportional hazards)

<table>
<thead>
<tr>
<th></th>
<th>Logrank</th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.65</td>
<td>0.53</td>
<td>0.42</td>
<td>0.35</td>
<td>0.29</td>
<td>0.21</td>
</tr>
</tbody>
</table>

- The power of the logrank test is very good.
- The difference of power between logrank and F-H is not so large.
Consider data generated under $q = 4$

<table>
<thead>
<tr>
<th></th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logrank</td>
<td>0.59</td>
<td>0.94</td>
<td>0.95</td>
<td>0.96</td>
<td>0.94</td>
</tr>
</tbody>
</table>

- The power of the F-H test is very good.
- The power of the logrank test is very low under non-proportional hazards assumption.
- The test is not sensitive to a variation of the value $q$.

⇒ Very reassuring for its use in clinical trials

Consider data generated under $q = 0$ (proportional hazards)

<table>
<thead>
<tr>
<th></th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logrank</td>
<td>0.65</td>
<td>0.42</td>
<td>0.35</td>
<td>0.29</td>
<td>0.21</td>
</tr>
</tbody>
</table>

- The power of the logrank test is very good.
- The difference of power between logrank and F-H is not so large.
Consider data generated under $q = 4$

<table>
<thead>
<tr>
<th></th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logrank</td>
<td>0.59</td>
<td>0.88</td>
<td>0.94</td>
<td>0.95</td>
<td><strong>0.96</strong></td>
</tr>
</tbody>
</table>

- The power of the F-H test is very good.
- The power of the logrank test is very low under non-proportional hazards assumption.
- The test is not sensitive to a variation of the value $q$.
  ⇒ Very reassuring for its use in clinical trials

Consider data generated under $q = 0$ (proportional hazards)

<table>
<thead>
<tr>
<th></th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logrank</td>
<td><strong>0.65</strong></td>
<td>0.53</td>
<td>0.42</td>
<td>0.35</td>
<td>0.29</td>
</tr>
</tbody>
</table>

- The power of the logrank test is very good.
- The difference of power between logrank and F-H is not so large.
Consider data generated under $q = 4$

<table>
<thead>
<tr>
<th></th>
<th>Logrank</th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.59</td>
<td>0.88</td>
<td>0.94</td>
<td>0.95</td>
<td><strong>0.96</strong></td>
<td>0.94</td>
</tr>
</tbody>
</table>

- The power of the F-H test is very good.
- The power of the logrank test is very low under non-proportional hazards assumption.
- The test is not sensitive to a variation of the value $q$.

⇒ Very reassuring for its use in clinical trials

Consider data generated under $q = 0$ (proportional hazards)

<table>
<thead>
<tr>
<th></th>
<th>Logrank</th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.65</td>
<td>0.53</td>
<td>0.42</td>
<td>0.35</td>
<td>0.29</td>
<td>0.21</td>
</tr>
</tbody>
</table>

- The power of the logrank test is very good.
- The difference of power between logrank and F-H is not so large.
Consider data generated under $q = 4$

<table>
<thead>
<tr>
<th></th>
<th>Logrank</th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.59</td>
<td>0.88</td>
<td>0.94</td>
<td>0.95</td>
<td><strong>0.96</strong></td>
<td>0.94</td>
<td></td>
</tr>
</tbody>
</table>

- The power of the F-H test is very good.
- The power of the logrank test is very low under non-proportional hazards assumption.
- The test is not sensitive to a variation of the value $q$.

⇒ Very reassuring for its use in clinical trials

Consider data generated under $q = 0$ (proportional hazards)

<table>
<thead>
<tr>
<th></th>
<th>Logrank</th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.65</td>
<td>0.53</td>
<td>0.42</td>
<td>0.35</td>
<td>0.29</td>
<td>0.21</td>
<td></td>
</tr>
</tbody>
</table>

- The power of the logrank test is very good.
- The difference of power between logrank and F-H is not so large.
Consider data generated under $q = 4$

<table>
<thead>
<tr>
<th></th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logrank</td>
<td>0.59</td>
<td>0.88</td>
<td>0.94</td>
<td>0.95</td>
<td>0.96</td>
</tr>
</tbody>
</table>

- The power of the F-H test is very good.
- The power of the logrank test is very low under non-proportional hazards assumption.
- The test is not sensitive to a variation of the value $q$.
  ⇒ Very reassuring for its use in clinical trials

Consider data generated under $q = 0$ (proportional hazards)

<table>
<thead>
<tr>
<th></th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logrank</td>
<td>0.65</td>
<td>0.53</td>
<td>0.42</td>
<td>0.35</td>
<td>0.29</td>
</tr>
</tbody>
</table>

- The power of the logrank test is very good.
- The difference of power between logrank and F-H is not so large.
Consider data generated under $q = 4$

<table>
<thead>
<tr>
<th></th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logrank</td>
<td>0.59</td>
<td>0.88</td>
<td>0.94</td>
<td>0.95</td>
<td>0.96</td>
</tr>
</tbody>
</table>

- The power of the F-H test is very good.
- The power of the logrank test is very low under non-proportional hazards assumption.
- The test is not sensitive to a variation of the value $q$.
  ⇒ Very reassuring for its use in clinical trials

Consider data generated under $q = 0$ (proportional hazards)

<table>
<thead>
<tr>
<th></th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logrank</td>
<td>0.65</td>
<td>0.53</td>
<td>0.42</td>
<td>0.35</td>
<td>0.29</td>
</tr>
</tbody>
</table>

- The power of the logrank test is very good.
- The difference of power between logrank and F-H is not so large.
Theorem

*In order to reach the level $\alpha$ with a power $\beta$ the sample size is*

$$n = \frac{2\sigma^2}{\mu^2} \left( z_{1-\alpha/2} + z_{\beta} \right)^2$$

*where*

$$\sigma^2 = \int_0^\tau (1 - S(s))^2 q \left( \frac{\pi^P(s)(\pi^T(s))^2}{(\pi(s))^2} d\Lambda^P(s) + \frac{(\pi^P(s))^2 \pi^T(s)}{(\pi(s))^2} d\Lambda^T(s) \right)$$

$$\mu = \int_0^\tau (1 - S(s))^q \frac{\pi^P(s)\pi^T(s)}{\pi(s)} (d\Lambda^P(s) - d\Lambda^T(s))$$

*with $S(s) = \frac{S^P(s) + S^T(s)}{2}$.*
The sample size increases when censoring increases,

decreases when the ratio between groups at the end of the study increases.

The sample size decreases with the value of parameter $q$ from $q = 1$ but can be more large than $q = 0$.

<table>
<thead>
<tr>
<th>Logrank</th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 2)$</th>
<th>$FH(q = 3)$</th>
<th>$FH(q = 4)$</th>
<th>$FH(q = 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2961</td>
<td>2332</td>
<td>1806</td>
<td>1474</td>
<td>1253</td>
<td>754</td>
</tr>
</tbody>
</table>
Theorem

Consider the placebo arm exponentially distributed. For $q > 1$ the function $\lambda^T$ defined by (2) has a single inflexion point.
Variation of hazard increases just before $t^*$ and decreases after $t^*$. It is in the neighbourhood of this time that we collect the best information.

Given a value of $q$, we are able to calculate the value of $t^*$.

Consider $n = 2000$, $\tau = 5$ years, $S^P(5) = 0.80$, $r = \frac{S^T(5) - S^P(5)}{1 - S^P(5)} = 20\%$.

<table>
<thead>
<tr>
<th>Logrank</th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 1.2)$</th>
<th>$FH(q = 1.4)$</th>
<th>$FH(q = 1.6)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>NA</td>
<td>0.95</td>
<td>1.65</td>
<td>2.25</td>
</tr>
<tr>
<td>$FH(q = 1.8)$</td>
<td>2.55</td>
<td>2.85</td>
<td>3.45</td>
<td></td>
</tr>
</tbody>
</table>
Inflexion point \( t^* \)

- Variation of hazard increases just before \( t^* \) and decreases after \( t^* \).

- It is in the neighbourhood of this time that we collect the best information.

- Given a value of \( q \), we are able to calculate the value of \( t^* \).

- Consider \( n = 2000, \tau = 5 \) years, \( S^p(5) = 0.80 \), \( r = \frac{S^T(5) - S^p(5)}{1 - S^p(5)} = 20\% \)

<table>
<thead>
<tr>
<th>Logrank</th>
<th>( FH(q = 1) )</th>
<th>( FH(q = 1.2) )</th>
<th>( FH(q = 1.4) )</th>
<th>( FH(q = 1.6) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>NA</td>
<td>0.95</td>
<td>1.65</td>
<td>2.25</td>
</tr>
<tr>
<td>( FH(q = 1.8) )</td>
<td>2.55</td>
<td>2.85</td>
<td>3.45</td>
<td></td>
</tr>
</tbody>
</table>
Variation of hazard increases just before \( t^* \) and decreases after \( t^* \).

It is **in the neighbourhood of this time** that we collect the best information.

Given a value of \( q \), we are able to calculate the value of \( t^* \).

Consider \( n = 2000, \tau = 5 \) years, \( S^P(5) = 0.80 \), 

\[
 r = \frac{S^T(5) - S^P(5)}{1 - S^P(5)} = 20\%
\]

<table>
<thead>
<tr>
<th>Logrank</th>
<th>FH((q = 1))</th>
<th>FH((q = 1.2))</th>
<th>FH((q = 1.4))</th>
<th>FH((q = 1.6))</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>NA</td>
<td>0.95</td>
<td>1.65</td>
<td>2.25</td>
</tr>
<tr>
<td>FH((q = 1.8))</td>
<td>FH((q = 2))</td>
<td>FH((q = 3))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.55</td>
<td>2.85</td>
<td>3.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Variation of hazard increases just before $t^*$ and decreases after $t^*$.

It is **in the neighbourhood of this time** that we collect the best information.

Given a value of $q$, we are able to calculate the value of $t^*$.

Consider $n = 2000$, $\tau = 5$ years, $S^P(5) = 0.80$, $r = \frac{S^T(5) - S^P(5)}{1 - S^P(5)} = 20\%$

<table>
<thead>
<tr>
<th>Logrank</th>
<th>$FH(q = 1)$</th>
<th>$FH(q = 1.2)$</th>
<th>$FH(q = 1.4)$</th>
<th>$FH(q = 1.6)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>NA</td>
<td>0.95</td>
<td>1.65</td>
<td>2.25</td>
</tr>
<tr>
<td>$FH(q = 1.8)$</td>
<td>$FH(q = 2)$</td>
<td>$FH(q = 3)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.55</td>
<td>2.85</td>
<td>3.45</td>
<td></td>
</tr>
</tbody>
</table>
Summary of the results on the choice of $q$:

- $q$ not too large
  in order to not neglect early events
- $q$ not too small
  in order to minimize sample size
- $q$ such that the associated inflexion point $t^*$ is of interest
  $t^* < \tau$
  $t^*$ collect the most information.

- A rough estimation of $q$ is enough.
  The test is not very sensitive to this value.
Summary of the results on the choice of $q$:

- $q$ not too large
  - in order to not neglect early events
- $q$ not too small
  - in order to minimize sample size
- $q$ such that the associated inflexion point $t^*$ is of interest
  - $t^* < \tau$
  - $t^*$ collect the most information.

- A rough estimation of $q$ is enough.
  - The test is not very sensitive to this value.
Recommendations

Summary of the results on the choice of $q$:

- $q$ not too large
  in order to not neglect early events
- $q$ not too small
  in order to minimize sample size
- $q$ such that the associated inflexion point $t^*$ is of interest
  $t^* < \tau$
  $t^*$ collect the most information.

- A rough estimation of $q$ is enough.
  The test is not very sensitive to this value.
Summary of the results on the choice of $q$:

- $q$ not too large
  in order to not neglect early events
- $q$ not too small
  in order to minimize sample size
- $q$ such that the associated inflexion point $t^* \leq \tau$
  $t^*$ collect the most information.

- A rough estimation of $q$ is enough.
  The test is not very sensitive to this value.
Summary of the results on the choice of $q$:

- $q$ not too large
  in order to not neglect early events
- $q$ not too small
  in order to minimize sample size
- $q$ such that the associated inflexion point $t^*$ is of interest
  
  $t^* < \tau$
  
  $t^*$ collect the most information.

- **A rough estimation of $q$ is enough.**
  The test is not very sensitive to this value.
Thank you for your attention
Counting processes and survival analysis.

A generalized Wilcoxon test for comparing arbitrarily singly-censored samples.
Biometrika 52, 203–223.

Censoring and stochastic integrals 124, v+178.

A class of rank test procedures for censored survival data.
Biometrika 69, 553–566.

Asymptotically efficient rank invariant test procedures.

Linear rank tests with right censored data.
Biometrika 65, 167–179.

On distribution-free tests for equality of survival distributions.
Biometrika 64, 156–160.
Cambridge : Cambridge University Press.