# Detecting sparse and weak deviations of non-proportional hazard in survival analysis

**Alon Kipnis (Reichman University)**
with Ben Galili and Zohar Yakhini (Technion)

Session: Detection and Estimation of Sparse and Weak Signals
Joint Statistical Meeting
August 2023
Toronto

# Overview

- Existing methods for comparing the survival of two groups are generally ineffective against differences in **hazard** occurring in a **few time instances** when those instances are **unknown** to us in advance

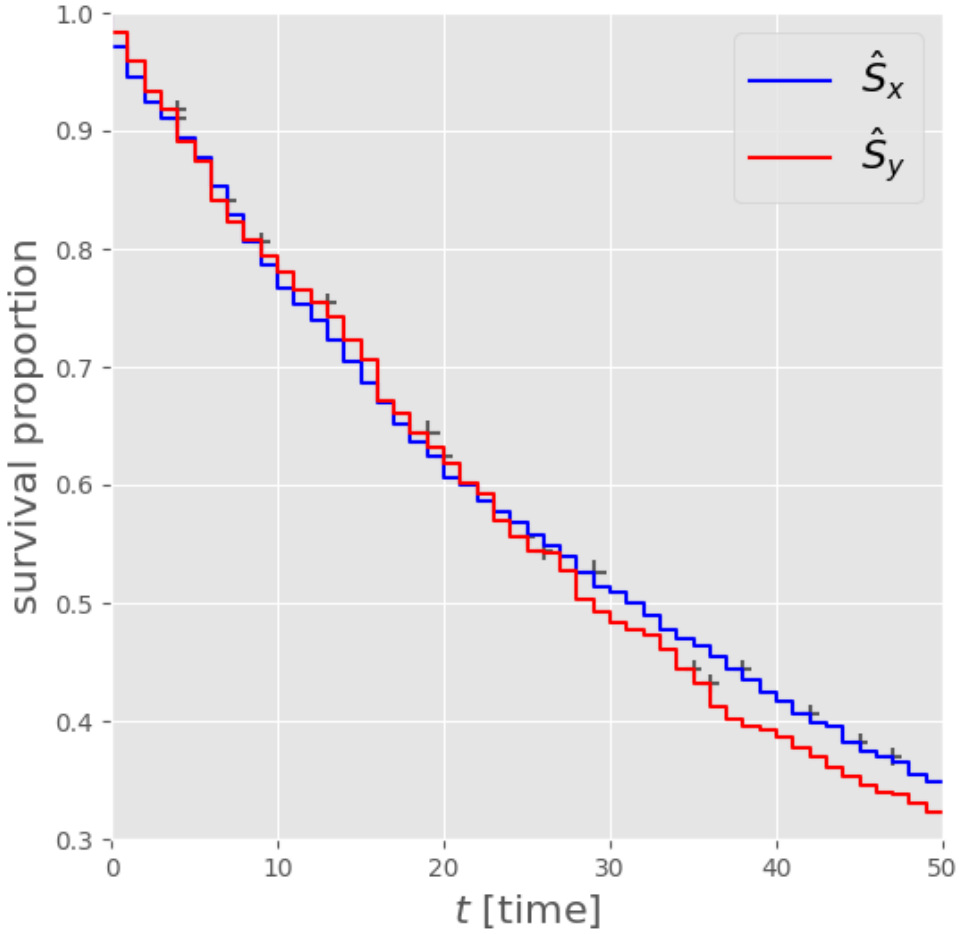
- We propose a method that is effective against such differences

# Survival Data and Analysis

| | Control ($X$) | | Treatment ($Y$) | |
|---|---|---|---|---|
| $t$ | at risk | events | at risk | events |
| 0 | 1000 | 29 | 1000 | 17 |
| 1 | 971 | 25 | 983 | 24 |
| 2 | 946 | 21 | 959 | 26 |
| 3 | 925 | 15 | 933 | 15 |
| 4 | 910 | 15 | 918 | 26 |
| 5 | 894 | 17 | 891 | 16 |
| 6 | 877 | 24 | 875 | 34 |
| 7 | 853 | 24 | 841 | 17 |
| 8 | 829 | 23 | 823 | 16 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 48 | 365 | 11 | 338 | 8 |
| 49 | 354 | 6 | 330 | 7 |
| 50 | 348 | | 323 | |

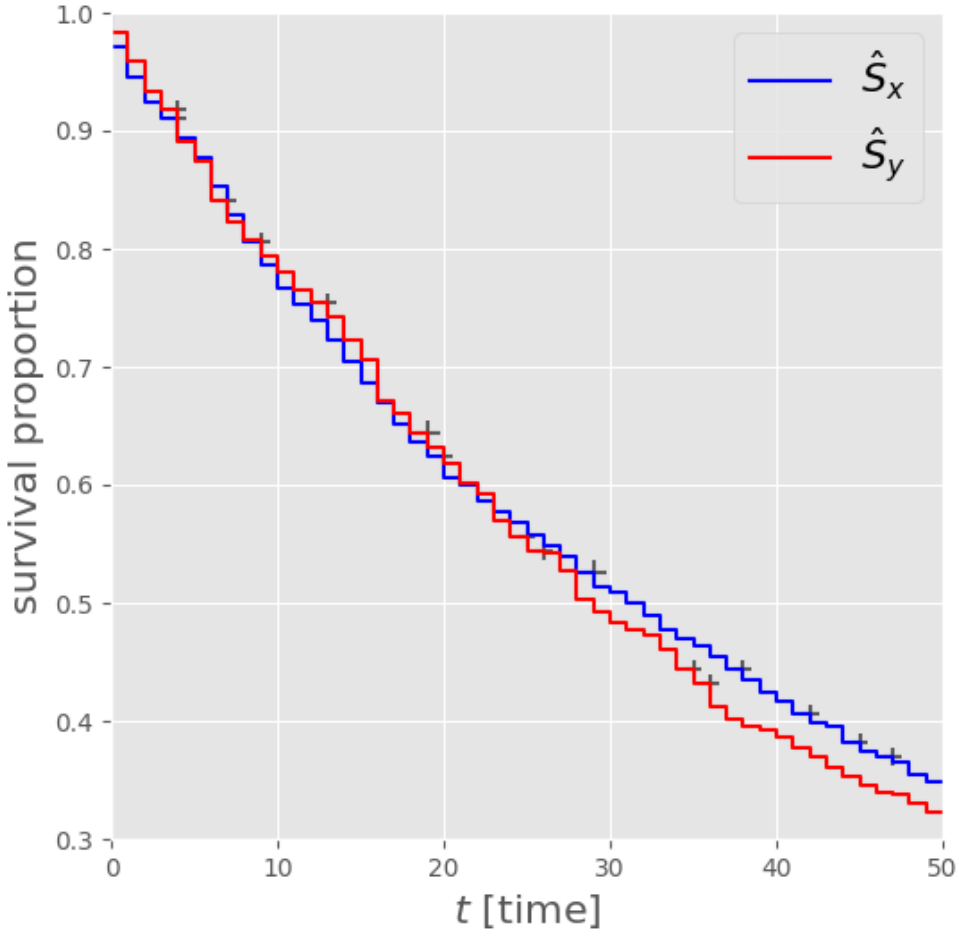
# Survival Data and Analysis

| $t$ | Control ($X$) at risk | Control ($X$) events | Treatment ($Y$) at risk | Treatment ($Y$) events |
|---|---|---|---|---|
| 0 | 1000 | 29 | 1000 | 17 |
| 1 | 971 | 25 | 983 | 24 |
| 2 | 946 | 21 | 959 | 26 |
| 3 | 925 | 15 | 933 | 15 |
| 4 | 910 | 15 | 918 | 26 |
| 5 | 894 | 17 | 891 | 16 |
| 6 | 877 | 24 | 875 | 34 |
| 7 | 853 | 24 | 841 | 17 |
| 8 | 829 | 23 | 823 | 16 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 48 | 365 | 11 | 338 | 8 |
| 49 | 354 | 6 | 330 | 7 |
| 50 | 348 | | 323 | |



# Survival Data and Analysis

| $t$ | Control ($X$) at risk | Control ($X$) events | Treatment ($Y$) at risk | Treatment ($Y$) events |
|---|---|---|---|---|
| 0 | 1000 | 29 | 1000 | 17 |
| 1 | 971 | 25 | 983 | 24 |
| 2 | 946 | 21 | 959 | 26 |
| 3 | 925 | 15 | 933 | 15 |
| 4 | 910 | 15 | 918 | 26 |
| 5 | 894 | 17 | 891 | 16 |
| 6 | 877 | 24 | 875 | 34 |
| 7 | 853 | 24 | 841 | 17 |
| 8 | 829 | 23 | 823 | 16 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 48 | 365 | 11 | 338 | 8 |
| 49 | 354 | 6 | 330 | 7 |
| 50 | 348 | | 323 | |




**The goal of the analysis:**

Determine whether the treatment has a non-null effect.

# The Log-Rank Test

- Mantel-Cox **log-rank test** [Mantel 1966]

## EVALUATION OF SURVIVAL DATA AND TWO NEW RANK ORDER STATISTICS ARISING IN ITS CONSIDERATION[1]

**Nathan Mantel**[2]

[PDF] Evaluation of survival data and two new rank order statistics arising in its consideration

N Mantel - Cancer Chemother Rep, 1966 - medicine.mcgill.ca

Survival-time patterns should be compared properly in their entirety rather than at isolated points only. Such overall comparison would require a value function for rating particular durations of survival, but no such function exists. A chi-square procedure is proposed for comparing two sets of life-table data in their entirety. The implicit value function for the pro cedure is reasonable in that it gives greater weight to earlier deaths. By considering the case in which the life-table intervals are arbitrarily short, it is seen to be essentially a rank order …

# The Log-Rank Test

| | Control ($X$) | | Treatment ($Y$) | |
|---|---|---|---|---|
| | at risk | events | at risk | events |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $t$ | $n_x(t-1)$ | $o_x(t)$ | $n_y(t-1)$ | $o_y(t)$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |

$\Rightarrow$

| total | Group $X$ | Group $Y$ |
|---|---|---|
| | | $\overbrace{\quad}^{K}$ |
| $M$ | $n_x(t-1)$ | $n_y(t-1)$ |
| $k$ | $o_x(t)$ | $o_y(t)$ |

$\Downarrow$

$$H_{0,t} : O_y(t) \sim \mathrm{HyG}(M, K, k)$$

# The Log-Rank Test

| | Control ($X$) | | Treatment ($Y$) | |
|---|---|---|---|---|
| | at risk | events | at risk | events |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $t$ | $n_x(t-1)$ | $o_x(t)$ | $n_y(t-1)$ | $o_y(t)$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |

$\Rightarrow$

| total | Group $X$ | Group $Y$ |
|---|---|---|
| | | $\overbrace{\quad}^{K}$ |
| $M$ | $n_x(t-1)$ | $n_y(t-1)$ |
| $k$ | $o_x(t)$ | $o_y(t)$ |

$$\Downarrow$$

$$H_{0,t} : O_y(t) \sim \mathrm{HyG}(M, K, k)$$

$$\mu(t) := \mathbb{E}\left[O_y(t)|H_{0,t}\right], \qquad V(t) = \mathrm{Var}\left[O_y(t)|H_{0,t}\right]$$

$$\mathrm{LR}_T := \frac{\sum_{t=1}^{T}\left(o_y(t) - \mu(t)\right)}{\sqrt{\sum_{t=1}^{T} V(t)}}$$

$$\mathrm{LR}_T \overset{D}{\approx} \mathcal{N}(0,1) \qquad \text{under the global null}$$

# The Log-Rank Test (cont'd)

- The **log-rank** test:

$$\mathrm{LR}_T := \frac{\sum_{t=1}^{T} (o_y(t) - \mu(t))}{\sqrt{\sum_{t=1}^{T} V(t)}}$$

# The Log-Rank Test (cont'd)

- The **log-rank** test:

$$\mathrm{LR}_T := \frac{\sum_{t=1}^{T} (o_y(t) - \mu(t))}{\sqrt{\sum_{t=1}^{T} V(t)}}$$

🙂 Can accommodate **censorship**

🙂 Asymptotically equivalent to the likelihood ratio test in a **proportional hazard** model [Breslow 1977]

☹ **Not sensitive to excessive hazard localized in time**

# The Log-Rank Test (cont'd)

- The **log-rank** test:

$$\mathrm{LR}_T := \frac{\sum_{t=1}^{T} (o_y(t) - \mu(t))}{\sqrt{\sum_{t=1}^{T} V(t)}}$$

  - 🙂 Can accommodate **censorship**
  - 🙂 Asymptotically equivalent to the likelihood ratio test in a **proportional hazard** model [Breslow 1977]
  - ☹ **Not sensitive to excessive hazard localized in time**
- Non-homogeneous Log-rank [Tarone & James 1977], [Lee 1996], [Liu et. al. 2022]
  - 🙂 Can be sensitive to non-proportional hazards, but
  - ☹ Not useful when **time instances** of excessive hazard are **apriori unknown**

# Survival Data with sparse and weak hazard departures

**Our goal:**

Attain sensitivity to **excessive hazard** localized in a **few** time instances such that we do not know in advance where those instances might be

# Survival Data with sparse and weak hazard departures

**Our goal:**

Attain sensitivity to **excessive hazard** localized in a **few** time instances such that we do not know in advance where those instances might be

Use cases:

- Identifying age-specific effects [Nuzhdin, Khazaeli, Curtsinger, 2005]
- Analyzing the effect of bursty “space weather” radiation on radioactive decay [Castro-Palacio et. al. 2020]

# Proposed method: Higher criticism of hypergeometric P-values (HCHG)

HCHG has two steps:

# Proposed method: Higher criticism of hypergeometric P-values (HCHG)

**HCHG** has two steps:

1. Many **exact hypergeometric** (HG) tests:

$$p_t := \Pr\left[\mathrm{HyG}(M, K, k) \geq o_y(t)\right], \qquad t = 1, \ldots, T,$$

$$M = n_x(t-1) + n_y(t-1), \quad K = n_y(t-1), \quad k = o_x(t) + o_y(t)$$

# Proposed method: Higher criticism of hypergeometric P-values (HCHG)

**HCHG** has two steps:

1. Many **exact hypergeometric** (HG) tests:

$$p_t := \Pr\left[\mathrm{HyG}(M, K, k) \geq o_y(t)\right], \qquad t = 1, \ldots, T,$$

$$M = n_x(t-1) + n_y(t-1), \quad K = n_y(t-1), \quad k = o_x(t) + o_y(t)$$

2. Global testing with **Higher Criticism** (HC):

$$\mathrm{HC} := \mathrm{HC}(p_1, \ldots, p_T) := \max_{i \leq T\gamma} \sqrt{T} \frac{i/T - p_{(i)}}{\sqrt{(i/T)\,(1 - i/T)}}$$

# Proposed method: Higher criticism of hypergeometric P-values (HCHG)

**HCHG** has two steps:

1. Many **exact hypergeometric** (HG) tests:

$$p_t := \Pr\left[\mathrm{HyG}(M, K, k) \geq o_y(t)\right], \qquad t = 1, \ldots, T,$$

$$M = n_x(t-1) + n_y(t-1), \quad K = n_y(t-1), \quad k = o_x(t) + o_y(t)$$

2. Global testing with **Higher Criticism** (HC):

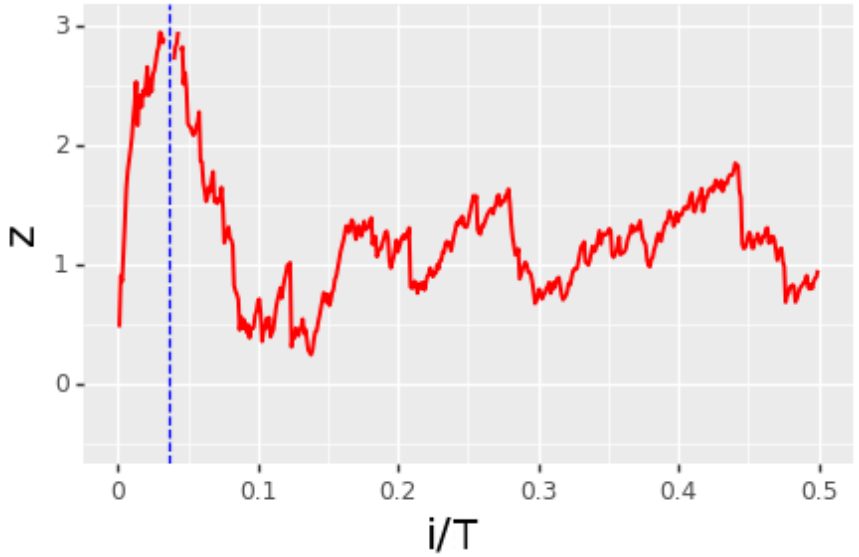
$$\mathrm{HC} := \mathrm{HC}(p_1, \ldots, p_T) := \max_{i \leq T\gamma} \sqrt{T} \frac{i/T - p_{(i)}}{\sqrt{(i/T)\,(1 - i/T)}}$$

Notable properties of $\mathrm{HCHG}$:

- Can accommodate censorship
- More sensitive than **log-rank** when instances of excessive risk are **sparse and weak** (later)
- Has a built-in mechanism to identify instances of excessive risk...

# Identifying instances of excessive risk

- Identifying instances of excessive risk via **HC thresholding** [Donoho & Jin 2008, 2009]:




$$\Delta^* = \left\{ t,\, p_t \leq p_{(i^\star)} \right\}, \qquad i^\star = \arg\max_{i \leq i_0} \sqrt{N} \frac{i/T - p_{(i)}}{\sqrt{(i/T)\,(1 - i/T)}}$$

# Analysis under Sparse and Weak Effect Model

- **Piece-wise exponential decay** model [Feigl & Zelen 1965], [Friedman 1982]:

$$N_x(0) = x_0, \qquad N_y(0) = y_0$$

for $t = 1, \ldots, T$

$$\begin{cases} O_x(t) \sim \text{Pois}(N_x(t-1)\bar{\lambda}_x(t)) \\ N_x(t) = N_x(t-1) - O_x(t) \end{cases} \qquad \begin{cases} O_y(t) \sim \text{Pois}(N_y(t-1)\bar{\lambda}_y(t)) \\ N_y(t) = N_y(t-1) - O_y(t) \end{cases}$$

# Analysis under Sparse and Weak Effect Model

- **Piece-wise exponential decay** model [Feigl & Zelen 1965], [Friedman 1982]:

$$N_x(0) = x_0, \qquad N_y(0) = y_0$$

for $t = 1, \ldots, T$

$$\begin{cases} O_x(t) \sim \text{Pois}(N_x(t-1)\bar{\lambda}_x(t)) \\ N_x(t) = N_x(t-1) - O_x(t) \end{cases} \qquad \begin{cases} O_y(t) \sim \text{Pois}(N_y(t-1)\bar{\lambda}_y(t)) \\ N_y(t) = N_y(t-1) - O_y(t) \end{cases}$$

- Non-homogeneous hazard alternative:

$$H_0 \ : \ \bar{\lambda}_x(t) = \bar{\lambda}_y(t), \qquad \forall t$$

$$H_1 \ : \ \bar{\lambda}_y(t) = \begin{cases} \bar{\lambda}_x(t) & \text{w.p. } 1-\epsilon \\ \left(\sqrt{\bar{\lambda}_x(t)} + \sqrt{\delta(t)}\right)^2 & \text{w.p. } \epsilon \end{cases}$$

# Analysis under Sparse and Weak Effect Model

- **Piece-wise exponential decay** model [Feigl & Zelen 1965], [Friedman 1982]:

$$N_x(0) = x_0, \qquad N_y(0) = y_0$$

for $t = 1, \ldots, T$

$$\begin{cases} O_x(t) \sim \text{Pois}(N_x(t-1)\bar{\lambda}_x(t)) \\ N_x(t) = N_x(t-1) - O_x(t) \end{cases} \qquad \begin{cases} O_y(t) \sim \text{Pois}(N_y(t-1)\bar{\lambda}_y(t)) \\ N_y(t) = N_y(t-1) - O_y(t) \end{cases}$$

- Non-homogeneous hazard alternative:

$$H_0 \ : \ \bar{\lambda}_x(t) = \bar{\lambda}_y(t), \qquad \forall t$$

$$H_1 \ : \ \bar{\lambda}_y(t) = \begin{cases} \bar{\lambda}_x(t) & \text{w.p. } 1-\epsilon \\ \left(\sqrt{\bar{\lambda}_x(t)} + \sqrt{\delta(t)}\right)^2 & \text{w.p. } \epsilon \end{cases}$$

- **Sparse** and **weak** calibration:
  - Individual effects are **sparse**: $\epsilon = T^{-\beta}, \qquad \beta \in (0,1)$
  - Individual effects are **weak**: $\delta(t)$ is small...

# Analysis under Sparse and Weak Effect Model (cont'd)

- Individual effects are **weak**:
  conditioned on $n_x(t-1)$, $n_y(t-1)$, $n_x(t)+o_y(t)$, hypergeometric
  P-values of **non-null** instances are **asymptotically**
  **log-chisquared** with a **moderate** location shift:

$$-2\log(p_t) \overset{D}{\approx} \left(\mathcal{N}(\sqrt{r\log(T)},1)\right)^2, \qquad r>0$$

  [Kipnis 2023], [Donoho & Kipnis 2023]

# Analysis under Sparse and Weak Effect Model (cont'd)

**Theorem**

HCHG *is asymptotically* powerful *if*

$$r > \rho(\beta) = \begin{cases} 2(\beta - 1/2) & \frac{1}{2} < \beta < \frac{3}{4}, \\ 2\left(1 - \sqrt{1-\beta}\right)^2 & \frac{3}{4} \leq \beta < 1, \end{cases}$$

*and asymptotically* powerless *if* $r < \rho(\beta)$

# Analysis under Sparse and Weak Effect Model (cont'd)

**Theorem**

HCHG *is asymptotically* **powerful** *if*

$$r > \rho(\beta) = \begin{cases} 2(\beta - 1/2) & \frac{1}{2} < \beta < \frac{3}{4}, \\ 2\left(1 - \sqrt{1-\beta}\right)^2 & \frac{3}{4} \leq \beta < 1, \end{cases}$$

*and asymptotically* **powerless** *if* $r < \rho(\beta)$

- $\rho(\beta)$ is the **two-sample sparse normal means** phase transition curve [Donoho & Kipnis 2023]. $\rho(\beta)/2$ is the **sparse normal means** phase transition curve [Ingster 1997], [Jin 2003], [Donoho & Jin 2004], [Mukherjee et. al 2015], [Arias-Castro & Wu 2015, 2018], [Jin & Ke 2016]...

# Analysis under Sparse and Weak Effect Model (cont'd)

**Theorem**

HCHG *is asymptotically* **powerful** *if*

$$r > \rho(\beta) = \begin{cases} 2(\beta - 1/2) & \frac{1}{2} < \beta < \frac{3}{4}, \\ 2\left(1 - \sqrt{1-\beta}\right)^2 & \frac{3}{4} \leq \beta < 1, \end{cases}$$

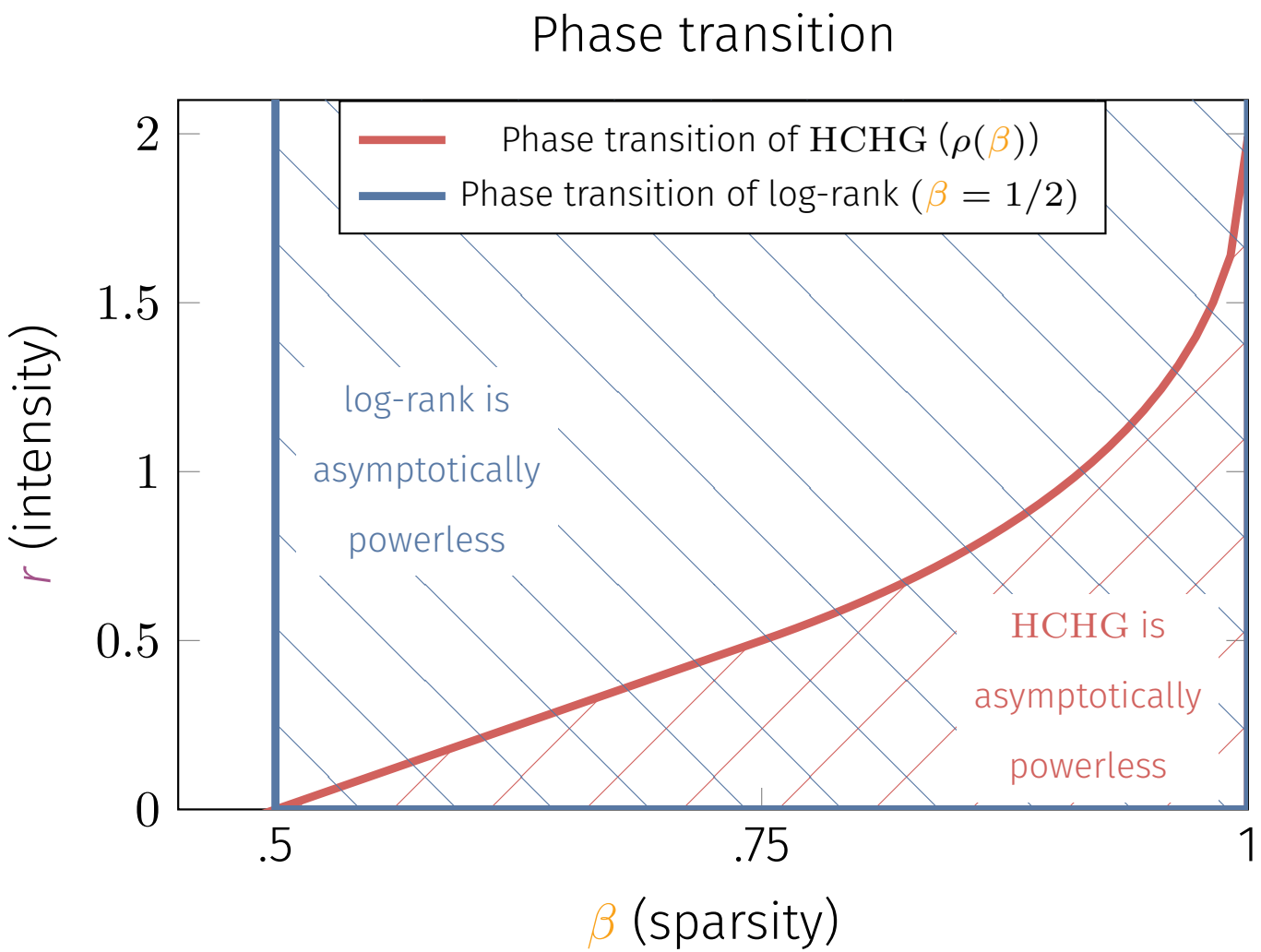
*and asymptotically* **powerless** *if* $r < \rho(\beta)$

- $\rho(\beta)$ is the **two-sample sparse normal means** phase transition curve [Donoho & Kipnis 2023]. $\rho(\beta)/2$ is the **sparse normal means** phase transition curve [Ingster 1997], [Jin 2003], [Donoho & Jin 2004], [Mukherjee et. al 2015], [Arias-Castro & Wu 2015, 2018], [Jin & Ke 2016]...

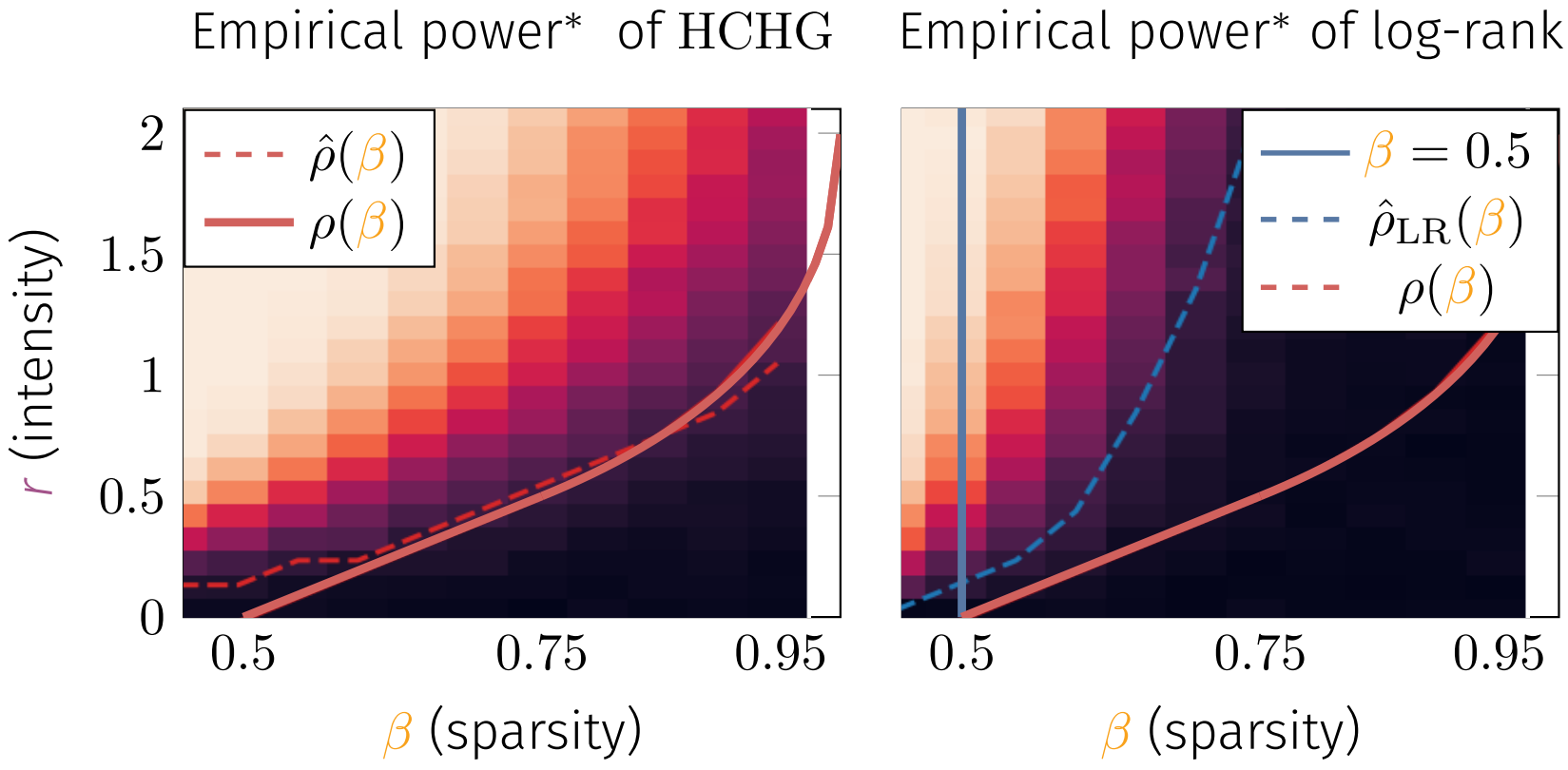
Also, when $\beta > 1/2$,

- **Log-rank** is asymptotically **powerless**
- **Fisher's combination** statistic of HG P-values is asymptotically **powerless**

# Asymptotic Power and Phase Transition

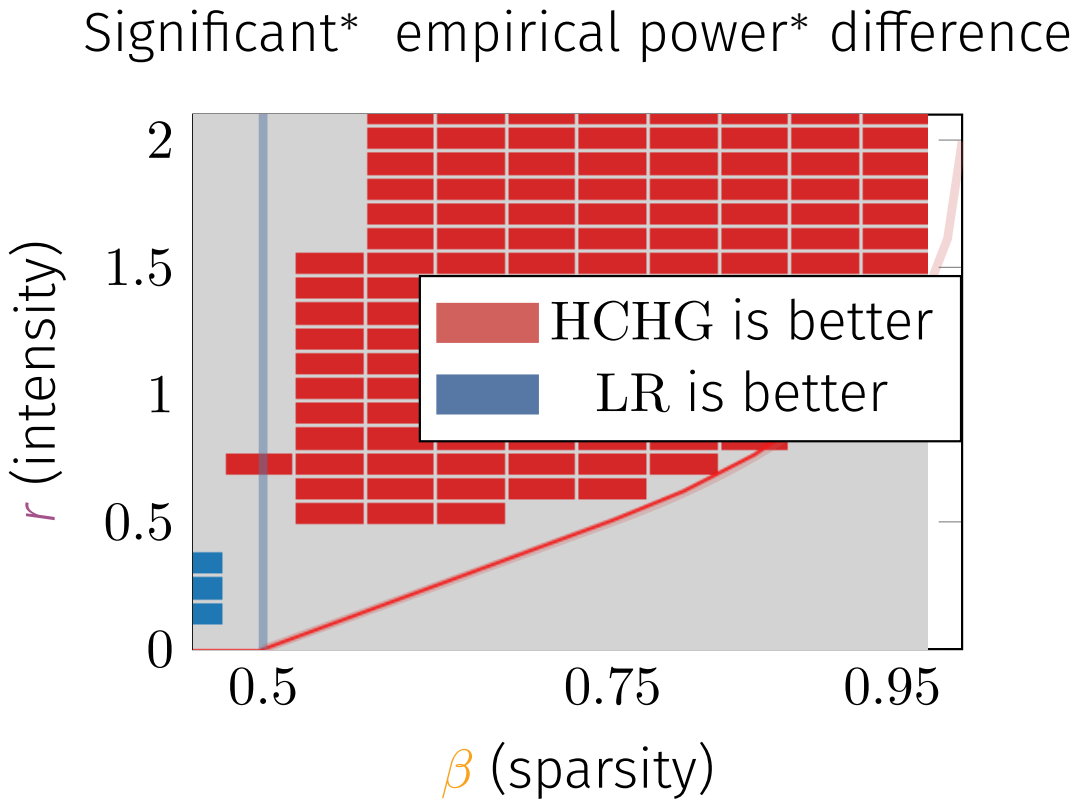


# Empirical Phase Transition




∗ at level $\alpha = 0.05$

# Empirical Power Comparison




∗ at level $\alpha = 0.05$

# Demonstration for Real Gene-Expression Data

- SCAN dataset [Saal et. al. 2015]:
  - $3,069$ breast cancer patients
  - Expression level of $8,702$ genes
- We partitioned each gene by its median expression level; yields $2 \times 8,702$ Control/Treatment assignments
- We consolidated events into $\approx 21$ day intervals (originally, $t \in [56, 2474]$ days)
- Example:

| Gene name: ADSS | | | | |
|---|---|---|---|---|
| | Below median ($X$) | | Above median ($Y$) | |
| $t$ | at risk | events | at risk | events |
| 0 | 1534 | 0 | 1535 | 1 |
| 1 | 1534 | 1 | 1534 | 1 |
| 2 | 1533 | 0 | 1533 | 0 |
| 3 | 1533 | 2 | 1533 | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 97 | 703 | 0 | 665 | 1 |
| 98 | 693 | 0 | 653 | 0 |
| 99 | 686 | 0 | 645 | 0 |

# Demonstration for Real Gene-Expression Data – Results

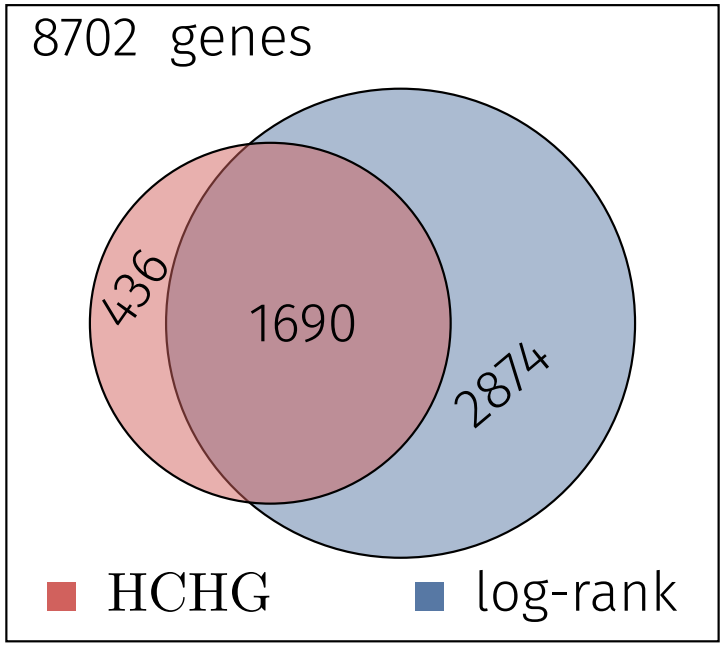



**Figure 1:** Number of genes with expression levels significantly[2] associated with survival according to HCHG(■) and log-rank (■)

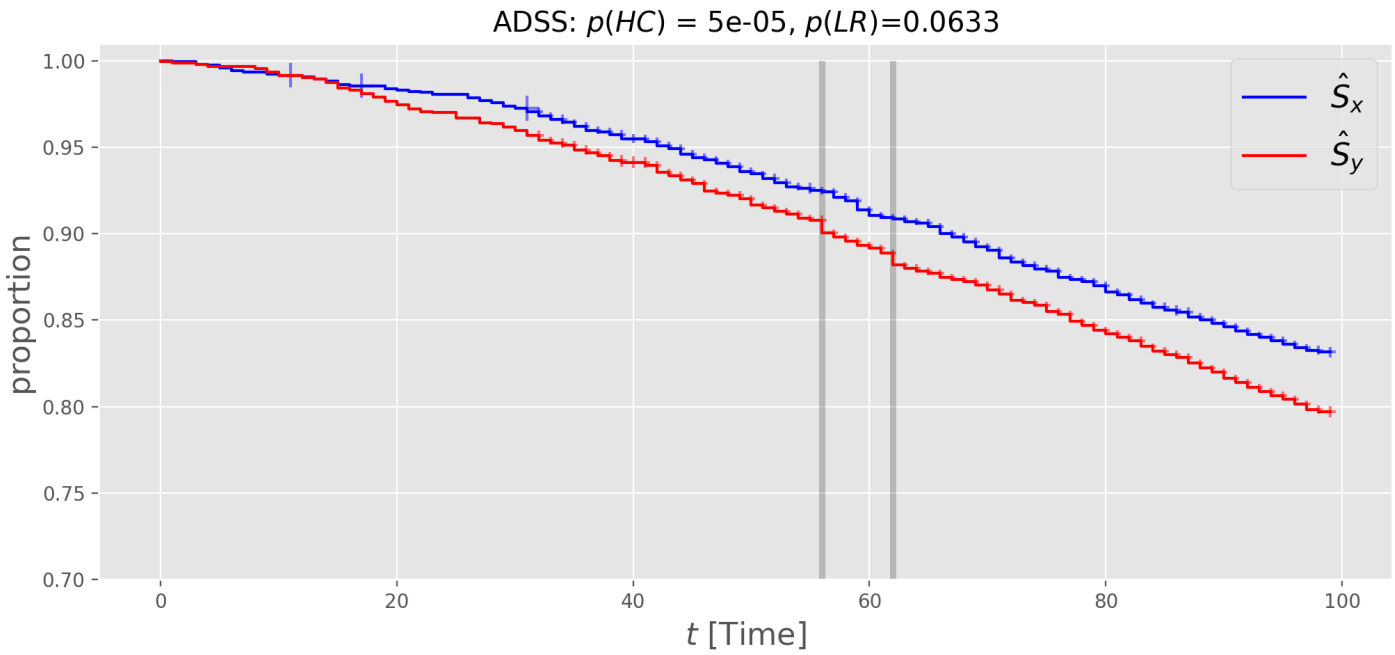
[1]at level $\alpha = 0.05$
[2]at level $\alpha = 0.05$

# Demonstration for Real Gene-Expression Data – Results (cont'd)

| Gene Name | HC (P-value) | Log-rank (P-value) | Increased Mortality |
|---|---|---|---|
| DCK | 0.00010 | 0.35505 | $>$ med |
| ADSS | 0.00005 | 0.0633 | $>$ med |
| KCTD9 | 0.01284 | 0.33369 | $>$ med |
| VAMP4 | 0.01271 | 0.20006 | $>$ med |
| TMEM38B | 0.02857 | 0.41772 | $<$ med |
| HIST1H3G | 0.02725 | 0.39828 | $<$ med |
| SIGMAR1 | 0.01180 | 0.16812 | $<$ med |
| POLDIP3 | 0.04683 | 0.33744 | $<$ med |
| SMG9 | 0.03775 | 0.22874 | $<$ med |
| FBXL12 | 0.01266 | 0.05641 | $>$ med |
| BTNL8 | 0.03934 | 0.05110 | $<$ med |

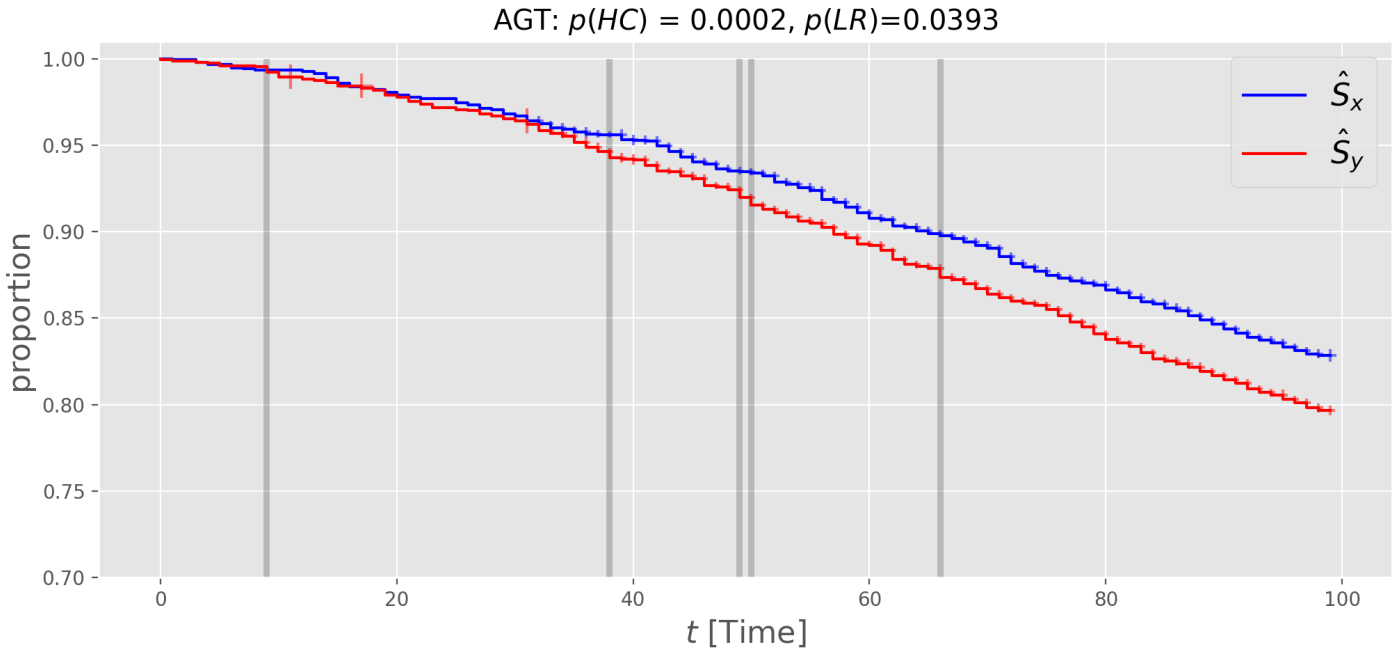
**Table 1:** Some genes in which **HCHG** identified a significantly lower survival rate in one group than the other while log-rank failed to do so.

# Demonstration for Real Gene-Expression Data – Results (cont'd)



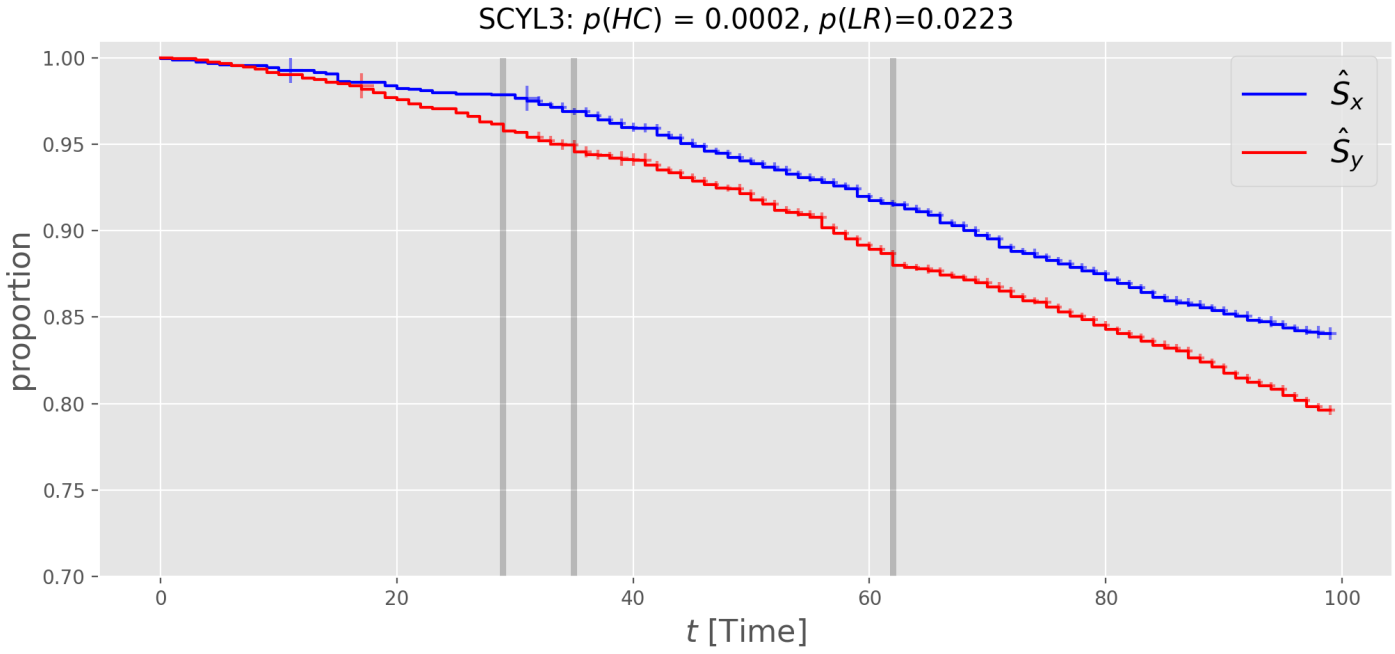

| $t$ | $n_x(t-1)$ | $n_y(t-1)$ | $o_x(t)$ | $o_y(t)$ | $p_t$ |
|---|---|---|---|---|---|
| 56.00 | 1203.00 | 1187.00 | 0.00 | 8.00 | 0.00 |
| 62.00 | 1120.00 | 1101.00 | 0.00 | 7.00 | 0.01 |

# Demonstration for Real Gene-Expression Data – Results (cont'd)




| $t$ | $n_x(t-1)$ | $n_y(t-1)$ | $o_x(t)$ | $o_y(t)$ | $p_t$ |
|---|---|---|---|---|---|
| 9.00 | 1524.00 | 1528.00 | 0.00 | 5.00 | 0.03 |
| 38.00 | 1412.00 | 1404.00 | 0.00 | 5.00 | 0.03 |
| 49.00 | 1281.00 | 1275.00 | 0.00 | 5.00 | 0.03 |
| 50.00 | 1271.00 | 1261.00 | 0.00 | 5.00 | 0.03 |
| 66.00 | 1069.00 | 1041.00 | 0.00 | 5.00 | 0.03 |

# Demonstration for Real Gene-Expression Data – Results (cont'd)




| $t$ | $n_x(t-1)$ | $n_y(t-1)$ | $o_x(t)$ | $o_y(t)$ | $p_t$ |
|---|---|---|---|---|---|
| 29.00 | 1500.00 | 1475.00 | 0.00 | 6.00 | 0.01 |
| 35.00 | 1454.00 | 1439.00 | 0.00 | 6.00 | 0.02 |
| 62.00 | 1111.00 | 1116.00 | 0.00 | 7.00 | 0.01 |

# Summary

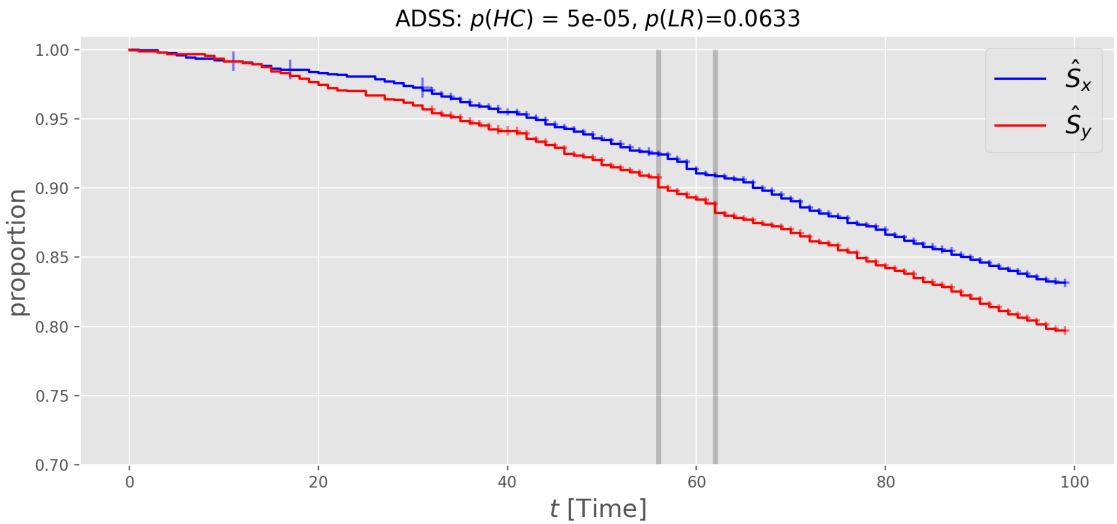
- HCHG is based on:
  1. Many exact hypergeometric tests
  2. Global testing with Higher Criticism

- HCHG is sensitive to **sparse and weak deviations** of non-proportional hazard
  - Theoretically: **more powerful** than existing methods in exponential decay **sparse and weak hazard departures setting**
  - Empirically: finds many discoveries not reported by the log-rank test

# References


- B. Galili, A. Kipnis and Z. Yakhini. (2023). Detecting rare and weak deviations of non-proportional hazard in survival analysis. *(on arxiv)*


The end.




| $t$ | $n_x(t-1)$ | $n_y(t-1)$ | $o_x(t)$ | $o_y(t)$ | $p_t$ |
|---|---|---|---|---|---|
| 56.00 | 1203.00 | 1187.00 | 0.00 | 8.00 | 0.00 |
| 62.00 | 1120.00 | 1101.00 | 0.00 | 7.00 | 0.01 |