MULTI-SAMPLE NONPARAMETRIC TESTS FOR PANEL COUNT DATA

Xingqiu Zhao

Joint work with N. Balakrishnan

McMaster University

OUTLINE

1. **Introduction**
2. **Modeling and inference**
3. **Asymptotic results**
4. **Simulation studies**
5. **Applications**
6. **Concluding remarks**

1 Introduction

- **Example: A bladder tumor study conducted by the Veterans Administration Co-operative Urological Research Group (Andrews and Herzberg, 1985).**

Multiple Tumor Recurrence Data

Patient	Treatment	t_1	C_1	t_2	C_2	t_3	C_3	$\cdots$	t_8	C_8	t_9	C_9
1	1	1	0									
2	1	3	1	9	1	21	8					
3	1	12	2	16	3							
46	1	2	1	15	3	24	4	$\cdots$	49	1	52	1
47	1	2	1	8	3	12	6	$\cdots$	49	1		

- **A bladder tumor study (cont.)**
  - **Three treatments:**

    Placebo (47), Thiotepa (31) and Pyridoxine (38)
  - **Data:**

    Visit times

    Numbers of tumors between clinical visits

    (*panel count data*)
  - **Interest: to determine the effect of treatment on the frequency of tumor recurrence**

Introduction (cont.)

- **Panel Count data**
  - $N(t)$: **the total number of occurrences of the event up to time** t
  - $N(t)$ **is observed only at discrete observation times**
  - **The study involves** k **groups**
  - **Let** $\Lambda_l(t)$ **denote the mean function of** $N(t)$ **corresponding to the** l**th group for** $l = 1, \ldots, k$
  - **Goal: to test the hypothesis** $H_0 : \Lambda_1(t) = \cdots = \Lambda_k(t)$

Introduction (cont.)

- **Analysis of panel count data**
  - Sun and Kalbfleisch (1995)
  - Wellner and Zhang (2000)
  - Sun and Wei (2000) and Zhang (2002)
  - Sun and Fang (2003)
  - Park, Sun and Zhao (2007)
  - Zhang (2006)

Introduction (cont.)

- Three estimators
  - Isotonic regression estimator
  - Nonparametric maximum pseudo-likelihood estimator (NPMPLE)
  - Nonparametric maximum likelihood estimator (NPMLE)
- No nonparametric tests have been discussed in the literature for panel count data based on the NPMLE

2 Statistical methods

2.1 Estimation of mean function

- **Notation**
  - $N(t)$: **a non-homogeneous Poisson process**
  - $\Lambda_0(t) = E(N(t))$
  - K: **an integer-valued random variable**
  - $T_K = (T_{K,1}, \ldots, T_{K,K})$ **and** $N_K = (N(T_{K,1}), \ldots, N(T_{K,K}))$
  - $X = (K, T_K, N_K)$ **and** $\mathbf{X} = (X_1, \ldots, X_n)$
  - $s_1, \ldots, s_m$: **ordered distinct observation times**
  - w_l: **number of the observations made at** s_l
  - $\bar{N}_l$: **mean value of the observations made at** s_l

Estimation of the mean function (cont.)

- **Isotonic regression approach**

The isotonic regression estimator $\hat{\Lambda}_n^{is}(t)$ is defined as the nondecreasing step function with possible jumps at s_ℓ's that minimizes

$$\sum_{\ell=1}^{m} w_\ell \{\bar{N}_\ell - \Lambda(s_\ell)\}^2 \tag{1}$$

subject to the order restriction $\Lambda(s_1) \leq \cdots \leq \Lambda(s_m)$.

$$\hat{\Lambda}_n^{is}(s_\ell) = \max_{r \leq \ell} \min_{s \geq \ell} \frac{\sum_{v=r}^{s} w_v \bar{n}_v}{\sum_{v=r}^{s} w_v}, \tag{2}$$

$$\ell = 1, ..., m .$$

Sun and Kalbfleisch (1995)

Estimation of the mean function (cont.)

- **Nonparametric maximum pseudo-likelihood approach**

  – **Pseudo log-likelihood function for Λ is**

$$l_n^{ps}(\Lambda|\mathbf{X}) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\{N_i(T_{K_i,j})\log(\Lambda(T_{K_i,j})) - \Lambda(T_{K_i,j-1})\} \quad (3)$$

  after omitting the parts independent of Λ.

  – **Define a nonparametric estimator $\hat{\Lambda}_n^{ps}$ of Λ_0 as a nondecreasing step function with possible jumps only occurring at $s_i, i = 1, \ldots, m$, that maximizes $l_n^{ps}(\Lambda|\mathbf{X})$**

  – $\hat{\Lambda}_n^{ps} = \hat{\Lambda}_n^{is}$

Wellner and Zhang (2000)

Estimation of the mean function (cont.)

- **Nonparametric maximum likelihood approach**
  - **The log-likelihood function for Λ is**

$$l_n(\Lambda|\mathbf{X}) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left(N_i(T_{K_i,j}) - N_i(T_{K_i,j-1})\right)\log\left(\Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j-1})\right) - \sum_{i=1}^{n}\Lambda(T_{K_i,K_i}) \tag{4}$$

    after omitting the parts independent of Λ.

  - **Define a nonparametric estimator $\hat{\Lambda}_n$ of Λ_0 as a nondecreasing step function with possible jumps only occurring at $s_i, i = 1, \ldots, m$, that maximizes $l_n(\Lambda|\mathbf{X})$**

2.2 Test statistics

n **independent subjects,** n_l **in group** l**,** $l = 1, \ldots, k$

$N_i(t)$**: counting process arising from subject** i

$(T_{K_i,1}, \ldots, T_{K_i,K_i})$**: observation times for subject** i

$\Lambda_l(t)$**: mean function of group** l

$H_0 : \Lambda_1 = \cdots = \Lambda_k = \Lambda_0$

- **Test statistics based on NPMPLE**

  - **For** $k = 2$

$$U_n^{ps} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{K_i} W_n(T_{K_i,j}) \{\hat{\Lambda}_{n_1}^{ps}(T_{K_i,j}) - \hat{\Lambda}_{n_2}^{ps}(T_{K_i,j})\} \tag{5}$$

  - **For** $k \geq 2$ **and** $l = 1, \ldots, k$**,**

$$U_n^{(ps,l)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{K_i} W_n^{(l)}(T_{K_i,j}) \{\hat{\Lambda}_n^{ps}(T_{K_i,j}) - \hat{\Lambda}_{n_l}^{ps}(T_{K_i,j})\} \tag{6}$$

  where $\hat{\Lambda}_{n_l}^{ps}$ **is the NPMPLE of** Λ_l **based on samples from subjects in group** l**, and** $W_n^{(l)}(t)$ **is a bounded weight process.**

- **Test Statistics based on NPMLE**

  – **For** $k \geq 2$ **and** $l = 1, \ldots, k$**,**

$$
\begin{aligned}
U_n^{(l)} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\sum_{j=1}^{K_i - 1} W_n^{(l)}(T_{K_i,j}) \hat{\Lambda}_n(T_{K_i,j}) \right. \\
&\qquad \times \left\{ \frac{\Delta \hat{\Lambda}_{n_l}(T_{K_i,j+1})}{\Delta \hat{\Lambda}_n(T_{K_i,j+1})} - \frac{\Delta \hat{\Lambda}_{n_l}(T_{K_i,j})}{\Delta \hat{\Lambda}_n(T_{K_i,j})} \right\} \\
&\qquad \left. + W_n^{(l)}(T_{K_i,K_i}) \hat{\Lambda}_n(T_{K_i,K_i}) \left\{ 1 - \frac{\Delta \hat{\Lambda}_{n_l}(T_{K_i,K_i})}{\Delta \hat{\Lambda}_n(T_{K_i,K_i})} \right\} \right] \quad (7)
\end{aligned}
$$

$$
\left(U_n^{(ps,l)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\sum_{j=1}^{K_i} W_n^{(l)}(T_{K_i,j}) \hat{\Lambda}_n(T_{K_i,j}) \left\{ 1 - \frac{\hat{\Lambda}_{n_l}(T_{K_i,j})}{\hat{\Lambda}_n(T_{K_i,j})} \right\} \right] \right)
$$

Test Statistics based on NPMLE (cont.)

– **For** $k \geq 2$ **and** $l = 2, \ldots, k$**,**

$$
\begin{aligned}
V_n^{(l)} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Bigg[\sum_{j=1}^{K_i-1} W_n^{(l)}(T_{K_i,j}) \hat{\Lambda}_n(T_{K_i,j}) \\
&\qquad \times \Bigg\{ \left(\frac{\Delta\hat{\Lambda}_{n_1}(T_{K_i,j+1})}{\Delta\hat{\Lambda}_n(T_{K_i,j+1})} - \frac{\Delta\hat{\Lambda}_{n_1}(T_{K_i,j})}{\Delta\hat{\Lambda}_n(T_{K_i,j})} \right) \\
&\qquad\quad - \left(\frac{\Delta\hat{\Lambda}_{n_l}(T_{K_i,j+1})}{\Delta\hat{\Lambda}_n(T_{K_i,j+1})} - \frac{\Delta\hat{\Lambda}_{n_l}(T_{K_i,j})}{\Delta\hat{\Lambda}_n(T_{K_i,j})} \right) \Bigg\} \\
&\qquad + W_n^{(l)}(T_{K_i,K_i}) \hat{\Lambda}_n(T_{K_i,K_i}) \left\{ \frac{\Delta\hat{\Lambda}_{n_l}(T_{K_i,K_i})}{\Delta\hat{\Lambda}_n(T_{K_i,K_i})} - \frac{\Delta\hat{\Lambda}_{n_1}(T_{K_i,K_i})}{\Delta\hat{\Lambda}_n(T_{K_i,K_i})} \right\} \Bigg]
\end{aligned}
\tag{8}
$$

Selection of weight process $W_n(t)$

- $W_n^{(1)}(t) = 1$

- $W_n^{(2)}(t) = Y_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(t \leq T_{K_i,K_i})$

- $W_n^{(3,l)}(t) = g(Y_{n_l}(t), Y_n(t))$ **where** $Y_{n_l}(t) = \frac{1}{n_l}\sum_{i\in S_l} I(t \leq T_{K_i,K_i})$

  where $S_l =$ **set of indices for subjects in group** l**.**

3 Asymptotic results

Result 1: Asymptotic normality of functional of $\hat{\Lambda}_n^{ps}$

Let $G(t) = E\left[\sum_{j=1}^{K} 1_{\{T_{K,j} \leq t\}}\right]$. *Then under mild conditions,*

$$\sqrt{n} \int_0^{\tau} W(t) \{\hat{\Lambda}_n^{ps}(t) - \Lambda_0(t)\} \, dG(t) \longrightarrow U_w \tag{9}$$

in distribution, where U_w *has a normal distribution with mean zero and variance that can be consistently estimated by*

$$\hat{\sigma}_w^2 = \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{K_i} W(T_{K_i,j}) \left\{ N_i(T_{K_i,j}) - \hat{\Lambda}_n^{ps}(T_{K_i,j}) \right\} \right]^2 \tag{10}$$

Result 2: Asymptotic distribution of $\mathbf{U}_n^{ps} = (U_n^{(ps,1)}, \cdots, U_n^{(ps,k)})^T$

Under H_0 *and mild conditions,* $\mathbf{U}_n^{ps}$ *has an asymptotic normal distribution with mean vector* $\mathbf{0}$ *and covariance matrix that can be consistently estimated by*

$$\hat{\mathbf{\Sigma}}_n^{ps} = \mathbf{\Gamma}_n \, \mathbf{diag}(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \cdots, \hat{\sigma}_k^2) \, \mathbf{\Gamma}_n', \tag{11}$$

where $\hat{\sigma}_l^2 = \frac{1}{n_l} \sum_{i \in S_l} \left[\sum_{j=1}^{K_i} W_n^{(l)}(T_{K_i,j}) \left\{ N_i(T_{K_i,j}) - \hat{\Lambda}_{n_l}^{ps}(T_{K_i,j}) \right\} \right]^2, \quad l = 1, \ldots, k,$ *and*

$$\mathbf{\Gamma}_n = \begin{pmatrix} \sqrt{\frac{n_1}{n}} - \sqrt{\frac{n}{n_1}} & \sqrt{\frac{n_2}{n}} & \cdots & \sqrt{\frac{n_k}{n}} \\ \sqrt{\frac{n_1}{n}} & \sqrt{\frac{n_2}{n}} - \sqrt{\frac{n}{n_2}} & \cdots & \sqrt{\frac{n_k}{n}} \\ \cdots & \cdots & \cdots & \cdots \\ \sqrt{\frac{n_1}{n}} & \sqrt{\frac{n_2}{n}} & \cdots & \sqrt{\frac{n_k}{n}} - \sqrt{\frac{n}{n_k}} \end{pmatrix}$$

Asymptotic results (cont.)

Similarly, we have shown

- **Asymptotic normality of functional of** $\hat{\Lambda}_n$
- **Asymptotic distribution of** $\mathbf{U}_n = (U_n^{(1)}, \cdots, U_n^{(k)})^T$
- **Asymptotic distribution of** $\mathbf{V}_n = (V_n^{(2)}, \cdots, V_n^{(k)})^T$

Nonparametric k-sample tests

- **Tests based on the NPMLE**
  - **Under** H_0, $T_n^{ps} = \mathbf{U}_0^{ps} \left(\hat{\mathbf{\Sigma}}_0^{ps} \right)^{-1} \mathbf{U}_0^{ps} \sim \chi^2(k-1)$, **where**

    $\mathbf{U}_0^{ps}$ = **the first** $(k-1)$ **components of** $\mathbf{U}_n^{ps}$

    $\hat{\mathbf{\Sigma}}_0^{ps}$ = **the matrix obtained by deleting the last row and column of** $\hat{\mathbf{\Sigma}}_n^{ps}$
  - T_{SF}^{ps}**: Sun and Fang (2003)**
  - T_{PSZ}^{ps}**: Park, Sun and Zhao (2007)**
  - T_Z^{ps}**: Zhang (2006)**

Nonparametric k-sample tests (cont.)

- **Tests based on the NPMLE**
  - Under H_0, $T_n^{(1)} = \mathbf{U}_0 \hat{\mathbf{\Sigma}}_0^{-1} \mathbf{U}_0 \sim \chi^2(k-1)$, where

    $\mathbf{U}_0$ = the first $(k-1)$ components of $\mathbf{U}_n$

    $\hat{\mathbf{\Sigma}}_0$ = the matrix obtained by deleting the last row and column of $\hat{\mathbf{\Sigma}}_{\mathbf{U}_n}$
  - Under H_0, $T_n^{(2)} = \mathbf{V}_n^T \hat{\mathbf{\Sigma}}_{\mathbf{V}_n}^{-1} \mathbf{V}_n \sim \chi^2(k-1)$
  - $T = V_n^{(2)}/\hat{\sigma} \sim N(0,1)$ for $k = 2$

4 Simulation studies

- **Generate k_i from the uniform distribution $U\{1, ..., 10\}$**
- **Given k_i, generate t_{ij}'s from $U\{1, ..., 10\}$**
- **N_i's are nonhomogeneous Poisson or mixed Poisson processes**
- **In particular, for given t_{ij}'s and random effect ν_i, suppose that $N_i(t_{ij})$ follows a Poisson distribution with mean $\Lambda_i(t) = \nu_i t$ for $i \in S_1$, $\Lambda_i(t) = \nu_i t \exp(\beta)$ for $i \in S_2$**
- **The results reported below are based on 1000 replications with $\nu_i = 1$ and $\nu_i \sim$ Gamma$(2, 1/2)$, respectively**

- **Choose the three weight processes:**

$$W_n^{(1)}(t) = 1$$

$$W_n^{(2)}(t) = Y_n(t) = \sum_{i=1}^{n} I(t \leq t_{i,k_i})/n$$

$$W_n^{(3)}(t) = Y_{n_1}(t)Y_{n_2}(t)/Y_n(t)$$

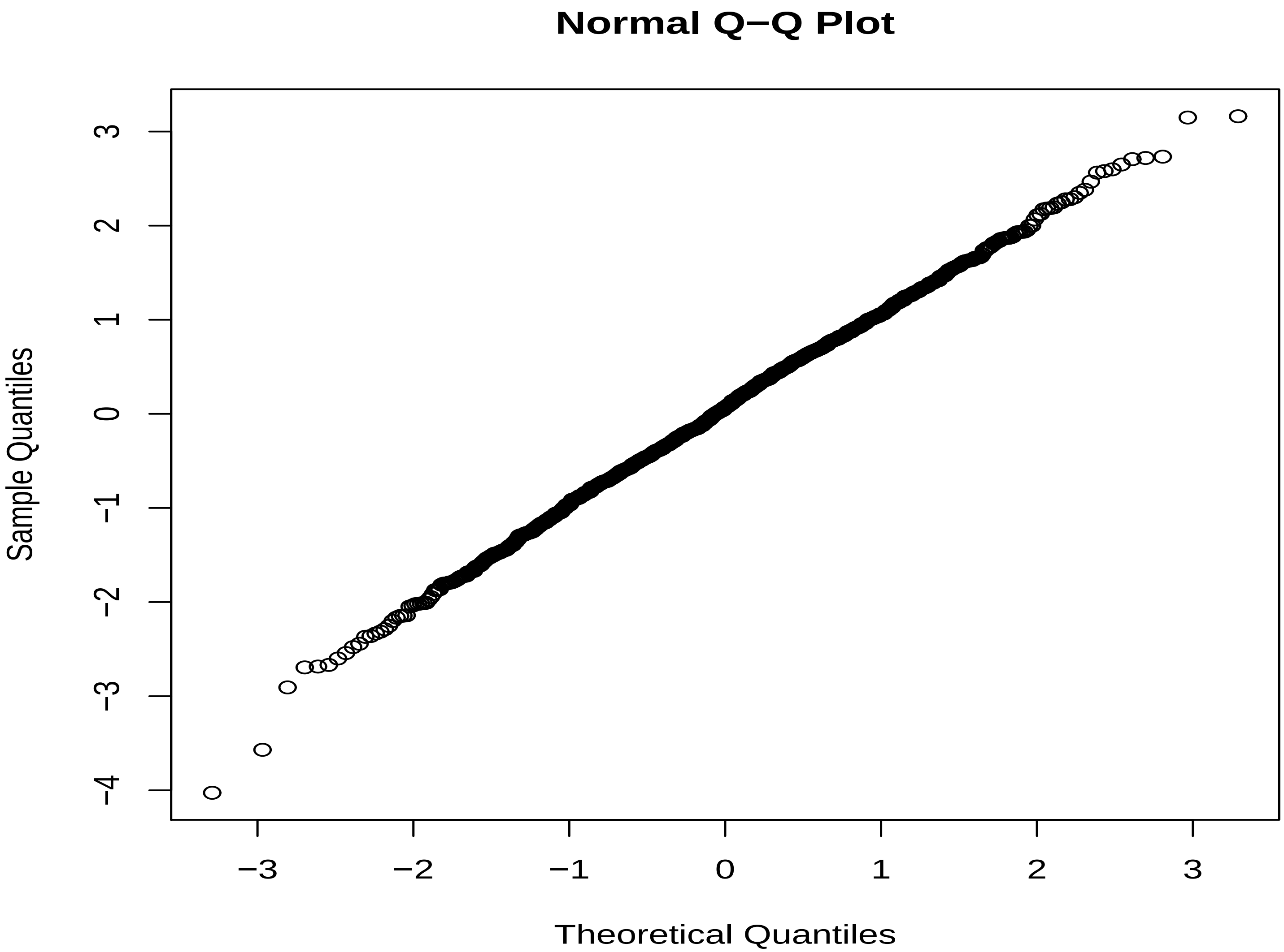

Figure 1: Simulation study. Normal quantile plot for T ($n = 100$).

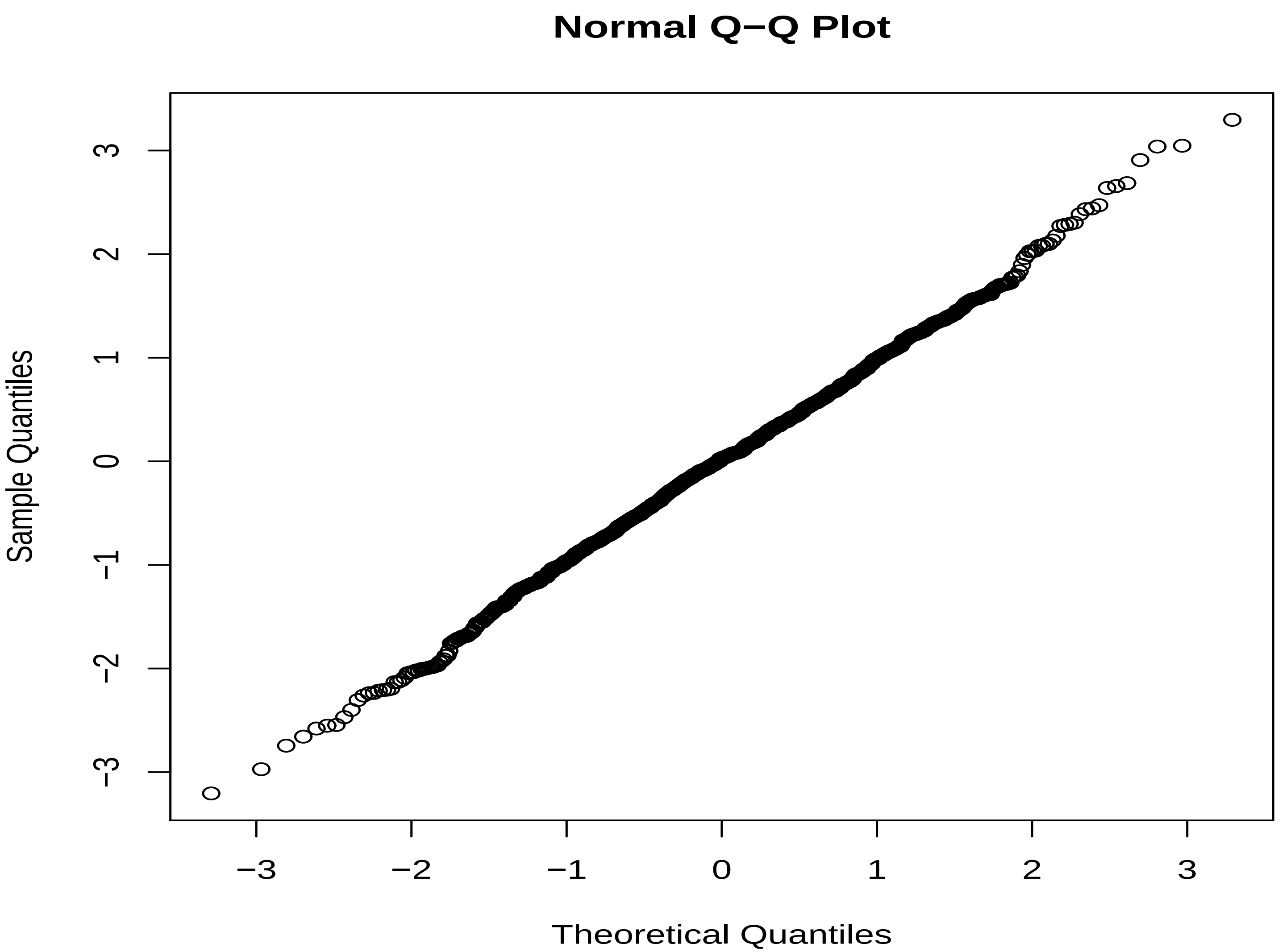

Figure 2: Simulation study. Normal quantile plot for T ($n = 200$).

Table 1: Estimated size and power of the tests for Poisson processes

β	T			T_{PSZ}^{ps} & T_{Z}^{ps}			T_{SF}^{ps}
	$W_n^{(1)}(t)$	$W_n^{(2)}(t)$	$W_n^{(3)}(t)$	$W_n^{(1)}(t)$	$W_n^{(2)}(t)$	$W_n^{(3)}(t)$	
				$n_1 = n_2 = 50$			
0.0	0.060	0.058	0.058	0.063	0.061	0.061	0.061
0.1	0.298	0.210	0.209	0.214	0.200	0.200	0.207
0.2	0.858	0.747	0.748	0.697	0.667	0.665	0.693
0.3	1.000	0.987	0.983	0.981	0.974	0.974	0.979
				$n_1 = n_2 = 100$			
0.0	0.047	0.047	0.047	0.044	0.046	0.046	0.043
0.1	0.542	0.472	0.471	0.423	0.405	0.405	0.422
0.2	0.993	0.967	0.964	0.958	0.948	0.947	0.950
0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 2: Estimated size and power of the tests for mixed Poisson processes

β	T			T_{PSZ}^{ps} & T_{Z}^{ps}			T_{SF}^{ps}
	$W_n^{(1)}(t)$	$W_n^{(2)}(t)$	$W_n^{(3)}(t)$	$W_n^{(1)}(t)$	$W_n^{(2)}(t)$	$W_n^{(3)}(t)$	
				$n_1 = n_2 = 50$			
0.0	0.043	0.040	0.042	0.037	0.040	0.040	0.035
0.1	0.100	0.097	0.097	0.084	0.085	0.085	0.083
0.2	0.221	0.205	0.207	0.185	0.184	0.184	0.183
0.3	0.458	0.407	0.408	0.380	0.375	0.375	0.370
				$n_1 = n_2 = 100$			
0.0	0.043	0.041	0.041	0.048	0.045	0.045	0.046
0.1	0.140	0.125	0.125	0.114	0.111	0.111	0.111
0.2	0.410	0.364	0.362	0.317	0.307	0.307	0.316
0.3	0.708	0.663	0.662	0.596	0.592	0.592	0.590

5 Illustrative examples

- **Example 1: A floating gallstones study** (Schonefield et al., 1981; Thall and Lachin, 1988)
  - Two groups: high dose (65) and placebo (48)
  - Data:

    Visit times

    Numbers of nausea between clinical visits
  - Interest: to compare the two treatments in terms of the incidence rates of nausea.

A floating gallstones study (cont.)

- T: p-value $= 0.837$ with $W_n^{(1)}$
- T: p-value < 0.01 with $W_n^{(2)}$
- T: p-value < 0.01 with $W_n^{(3)}$
- T_{PSZ} and T_Z: p-values $= 0.454, 0.417$ and 0.413 with three weights (Park et al., 2007)
- T_{SF}: p-value $= 0.1428$ (Sun and Fang, 2003)

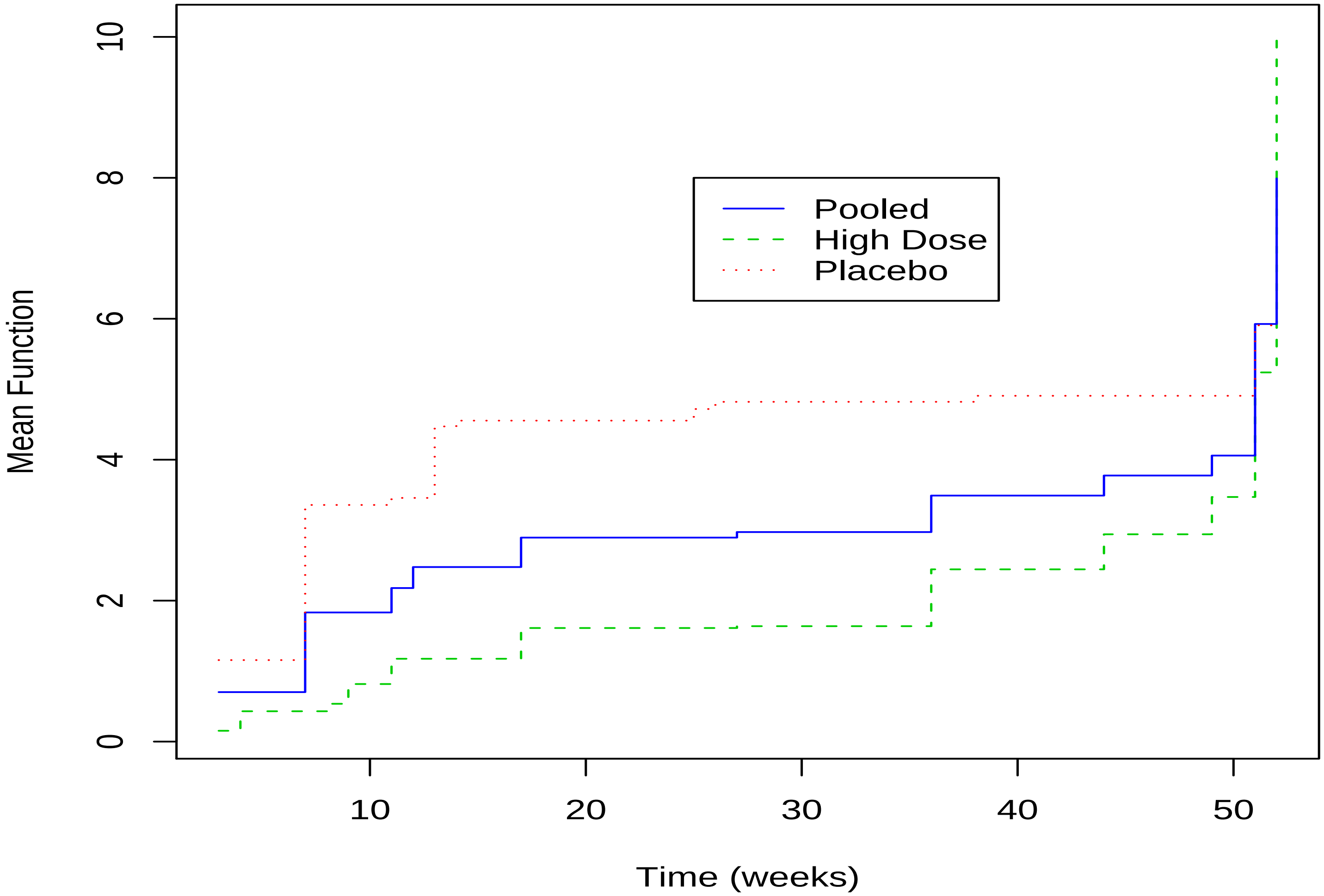

Figure 3: Floating gallstone study. NPMLEs of the mean functions.

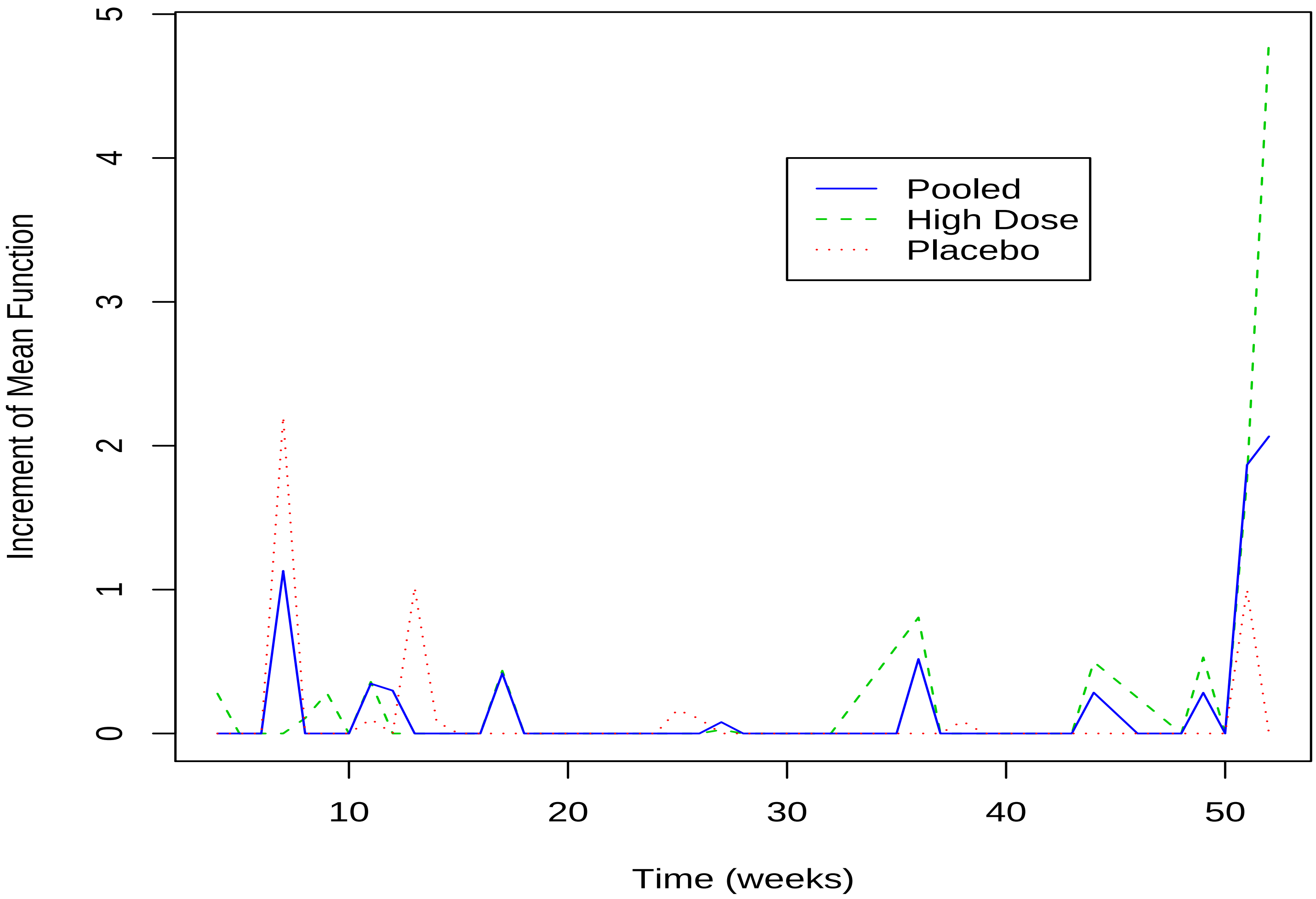

Figure 4: Floating gallstone study. Increments of the estimated mean functions.

- **Example 2: A bladder tumor study**
  - $T_n^{(2)}$: p-value $= 0.195$ **with** $W_n^{(l)}(t) = 1$
  - $T_n^{(2)}$: p-value < 0.01 **with** $W_n^{(l)}(t) = Y_n(t)$
  - $T_n^{(2)}$: p-value < 0.01 **with** $W_n^{(l)}(t) = 1 - Y_n(t)$
  - T_Z: p-**values: 0.0851, 0.1445 and 0.0840 with** $W_n^{(1)}$, $W_n^{(2)}$ **and** $W_n^{(3)}$ **(Zhang, 2006)**

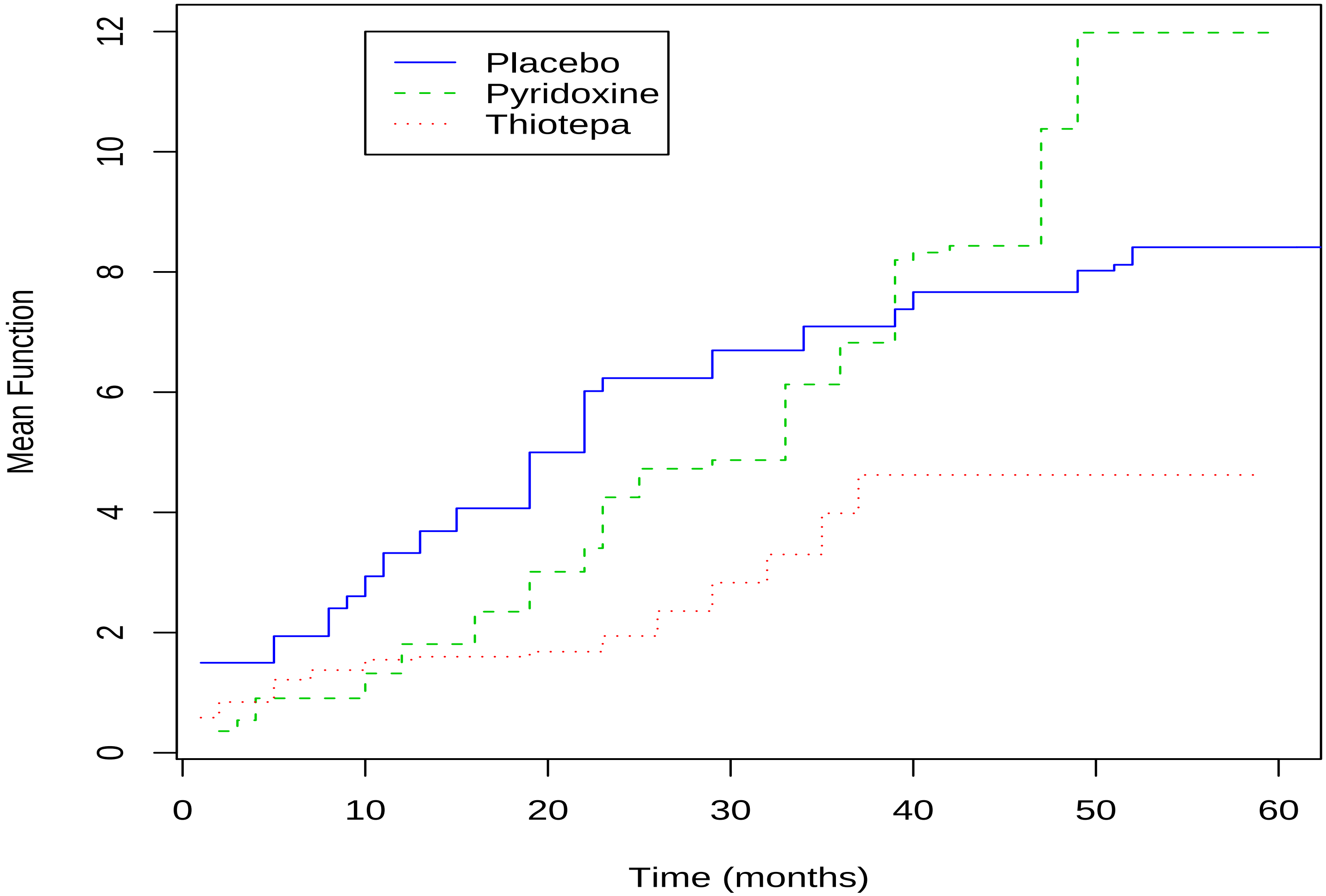

Figure 5: Bladder tumor study. NPMLEs of the mean functions.

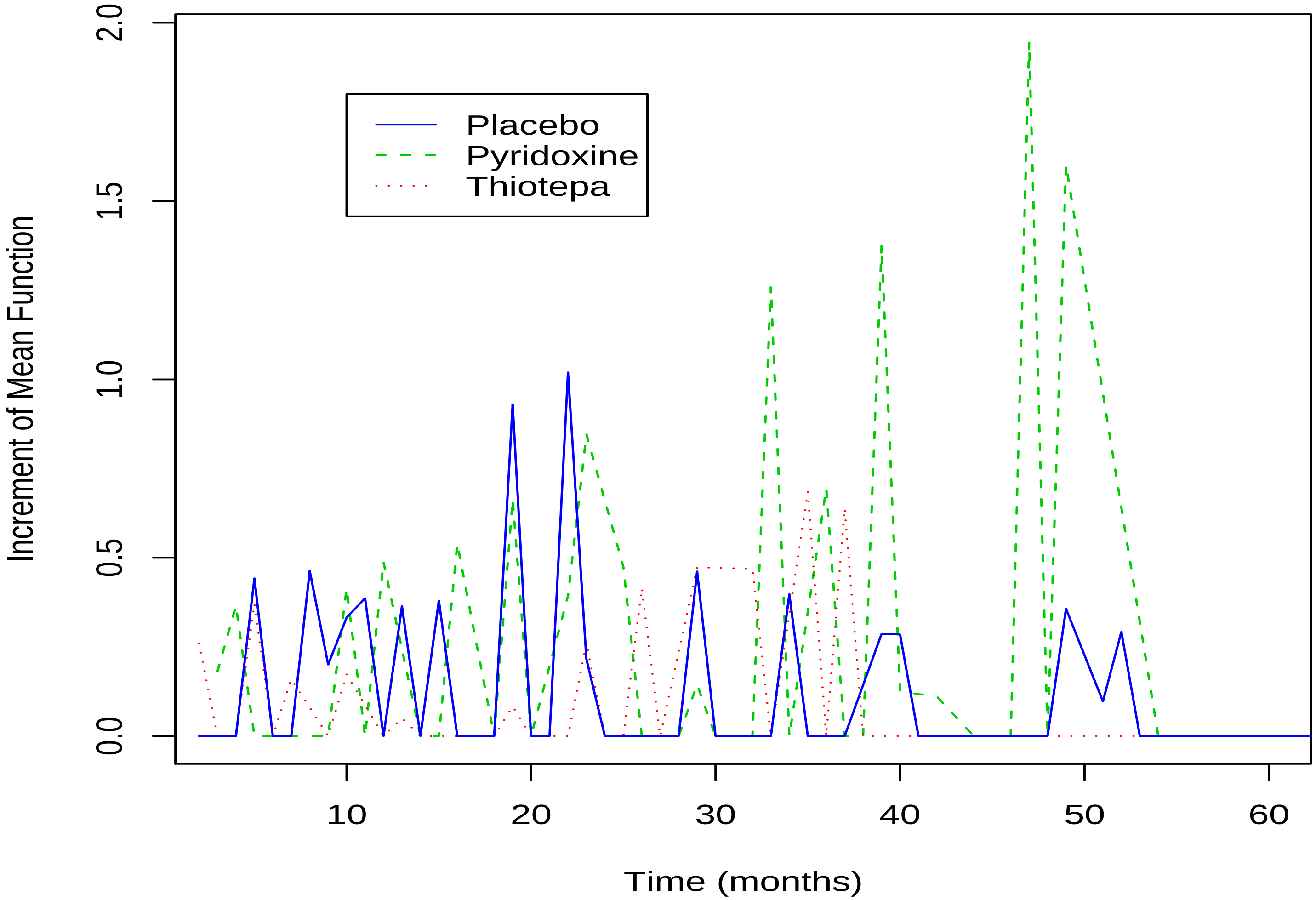

Figure 6: Bladder tumor study. Increments of the estimated mean functions.

6 Concluding remarks

- New nonparametric tests are developed for panel count data

- Simulation studies suggest that the proposed methods work quite well and the tests based on the NPMLE are more powerful than those based on NPMPLE

- The presented approach applies to more general situations than the existing methods

Future work

- **Study the properties of the test statistic under alternatives**
- **Generalize the proposed approach to situations where the underlying distribution of (K, T) may be different for different treatment groups**
- **Extend the proposed approach to situations where (K, T) and N may be dependent**

References

Balakrishnan, N. & Zhao, X. (2007). **A class of multi-sample nonparametric tests for panel count data.**

Balakrishnan, N. & Zhao, X. (2007). **New multi-sample nonparametric tests for panel count data.**

PARK, D-H., SUN, J. & ZHAO, X. (2007). **A class of two-sample nonparametric tests for panel count data.** *Comm. Statist. Theory Methods* **36, 1611–1625.**

SUN, J. & FANG, H. B. (2003). **A nonparametric test for panel count data.** *Biometrika* **90, 199-208.**

WELLNER, J. A. & ZHANG, Y. (2000). **Two estimators of the mean of a counting process with panel count data.** *Ann. Statist.* **28, 779-814.**

ZHANG, Y. (2006). **Nonparametric k-sample test with panel count data.** *Biometrika* **39, 777-790.**

THANK YOU