

## 2. Porównywanie krzywych przeżycia

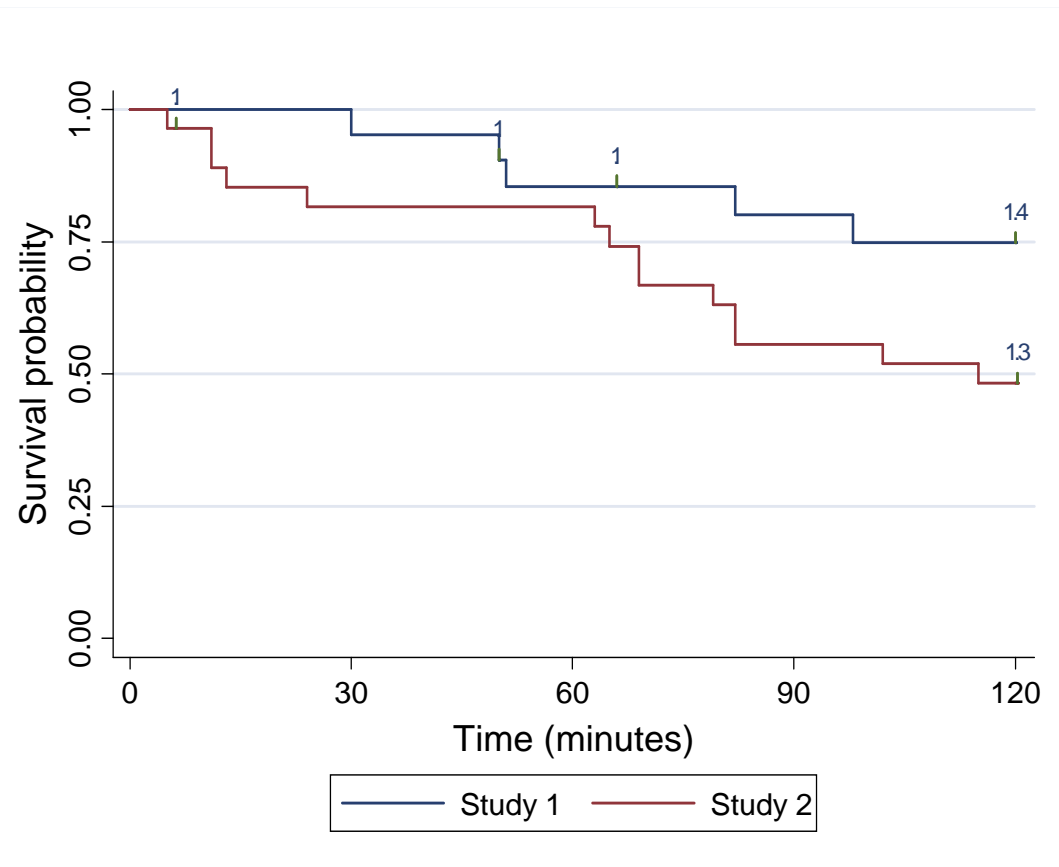
Test logrank

Modyfikacje

Inne testy

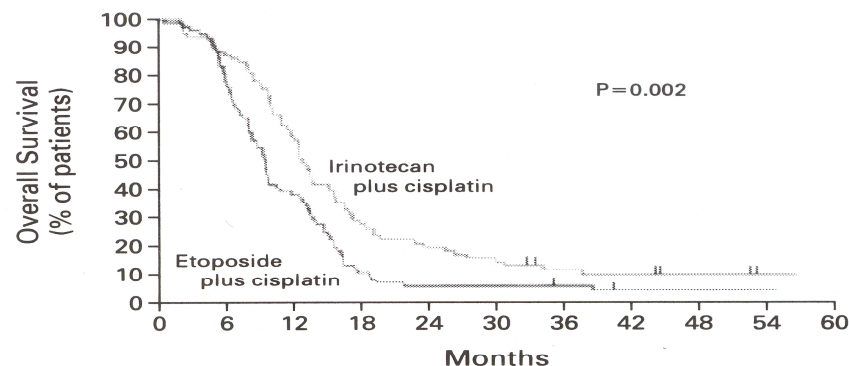
# Choroba lokomocyjna: różnice w krzywych przeżycia

- ◆ Dla ustalonej chwili czasu  $t$ , p-stwo, że osoba „przeżyła” bez torsji było mniejsze w drugim badaniu...
- ◆ ... czyli czas do torsji był średnio krótszy.
- ◆ Może jest to jedynie przypadkowa różnica?



# Próba kliniczna SCLC

Noda *et al.*,  
*NEJM* (2002)

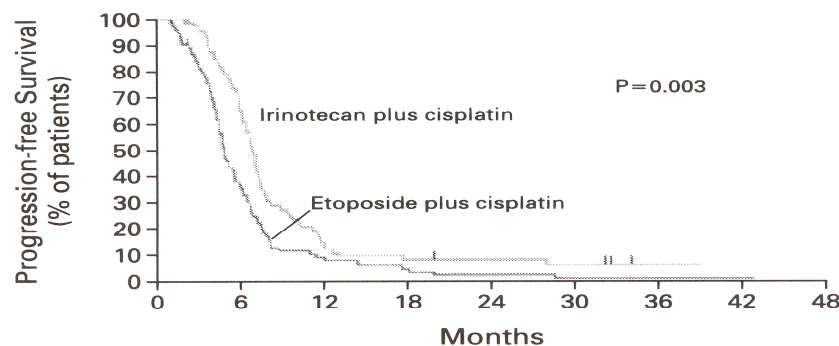


No. AT Risk

Irinotecan plus cisplatin	77	67	45	21	15	11	7
Etoposide plus cisplatin	77	60	29	8	4	4	3

**Figure 1.** Overall Survival of Patients with Extensive Small-Cell Lung Cancer Who Were Assigned to Treatment with Irinotecan plus Cisplatin or Etoposide plus Cisplatin.

The tick marks indicate patients whose data were censored.



No. AT Risk

Irinotecan plus cisplatin	77	47	9	9
Etoposide plus cisplatin	77	27	6	2

**Figure 2.** Progression-free Survival of Patients with Extensive Small-Cell Lung Cancer Who Were Assigned to Treatment with Irinotecan plus Cisplatin or Etoposide plus Cisplatin.

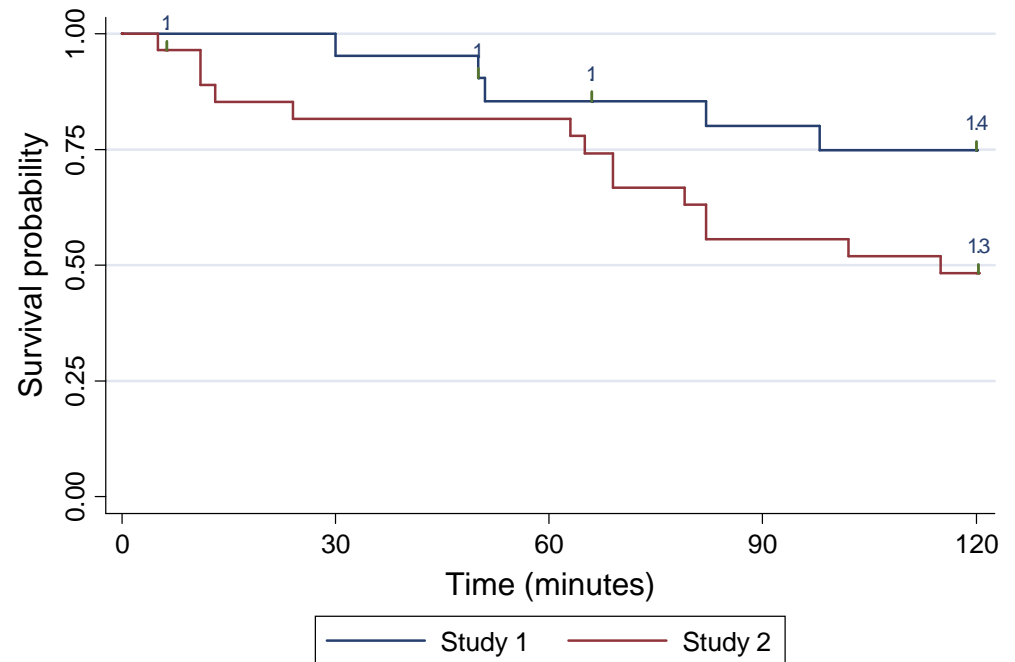
The tick marks indicate patients whose data were censored.

# Choroba lokomocyjna: różnice w krzywych przeżycia

♦ P-stwo przeżycia i 95% przedz. ufności (Greenwood):

$t$	1 <sup>sze</sup> badanie	2 <sup>gie</sup> badanie
30	.95 [0.86, 1.04]	.81 [0.67, 0.96]
60	.85 [0.70, 1.00]	.81 [0.67, 0.96]
90	.80 [0.63, 0.98]	.56 [0.37, 0.74]

♦ Mała różnica dla 60 min, większa dla 90...



# Porównywanie krzywych przeżycia

- ◆ Porównanie krzywych dla ustalonej chwili czasu wymaga wyboru tej chwili przed zgromadzeniem danych
- ◆ Powinno opierać się na błędzie standardowym  $\ln\{-\ln S(t)\}$ 
  - rozkład  $\ln\{-\ln S(t)\}$  jest bliższy normalnemu
- ◆ W przeciwnym przypadku, wynik jest problematyczny:
  - potencjalnie obciążony gdy chwila czasu wybrana a posteriori
  - wielokrotne porównania

# Dwupróbkowy test logrank

- ◆ Nieparametryczny
- ◆ *Hipoteza zerowa:  $S_1(t)=S_2(t)$*   
(funkcje przeżycia w obu grupach takie same)
- ◆ *Hipoteza alternatywna:  $S_1(t) \neq S_2(t)$*   
(funkcje się różnią)
- ◆ Idea: przy założeniu hipotezy zerowej, porównujemy oczekiwaną i obserwowaną liczbę zdarzeń

# Dwupróbkowy test logrank (1)

- ◆ Dla każdego czasu zdarzenia  $t_{(k)}$ , konstruujemy tabelę

Grupa	Zbiór ryzyka dla $t_{(k)}$	Zdarzenia obserwowane dla $t_{(k)}$	Zdarzenia oczekiwane dla $t_{(k)}$
1	$n_{1k}$	$d_{1k}$	$e_{1k} = d_k (n_{1k} / n_k)$
2	$n_{2k}$	$d_{2k}$	$e_{2k} = d_k (n_{2k} / n_k)$
Razem	$n_k$	$d_k$	$e_k = e_{1k} + e_{2k} = d_k$

- ◆ Przy założeniu hipotezy zerowej, łączny rozkład warunkowy (ze względu na zdarzenia i cenzurowanie do  $t_{(k)}$ )  $d_{1k}$  i  $d_{2k}$  to

$$\prod_{j=1}^2 \binom{n_{jk}}{d_{jk}} \lambda_k^{d_{jk}} (1 - \lambda_k)^{n_{jk} - d_{jk}}$$

## Dwupróbkowy test logrank (2)

- ◆ Dla każdego czasu zdarzenia  $t_{(k)}$ , konstruujemy tabelę

Grupa	Zbiór ryzyka dla $t_{(k)}$	Zdarzenia obserwowane dla $t_{(k)}$	Zdarzenia oczekiwane dla $t_{(k)}$
1	$n_{1k}$	$d_{1k}$	$e_{1k} = d_k (n_{1k} / n_k)$
2	$n_{2k}$	$d_{2k}$	$e_{2k} = d_k (n_{2k} / n_k)$
Razem	$n_k$	$d_k$	$e_k = e_{1k} + e_{2k} = d_k$

- ◆ Rozkład warunkowy  $d_{1k}$  i  $d_{2k}$ , przy danym  $d_k$ , jest rozkładem hipergeometrycznym:

$$\binom{n_k}{d_k}^{-1} \prod_{j=1}^2 \binom{n_{jk}}{d_{jk}}$$



## Dwupróbkowy test logrank (3)

- ◆ Dla każdego czasu zdarzenia  $t_{(k)}$ , konstruujemy tabelę

Grupa	Zbiór ryzyka dla $t_{(k)}$	Zdarzenia obserwowane dla $t_{(k)}$	Zdarzenia oczekiwane dla $t_{(k)}$
1	$n_{1k}$	$d_{1k}$	$e_{1k} = d_k (n_{1k} / n_k)$
2	$n_{2k}$	$d_{2k}$	$e_{2k} = d_k (n_{2k} / n_k)$
Razem	$n_k$	$d_k$	$e_k = e_{1k} + e_{2k} = d_k$

- ◆ Dla rozkładu hipergeometrycznego:

$$E(D_{jk}) = n_{jk} d_k / n_k$$

$$\text{Var}(D_{jk}) = n_{jk} (n_k - n_{jk}) d_k (n_k - d_k) / \{n_k^2 (n_k - 1)\}$$

$$\text{Cov}(D_{jk}, D_{j'k}) = -n_{jk} n_{j'k} d_k (n_k - d_k) / \{n_k^2 (n_k - 1)\}$$

## Dwupróbkowy test logrank (4)

- ◆ Dla każdego czasu zdarzenia  $t_{(k)}$ , konstruujemy tabelę

Grupa	Zbiór ryzyka dla $t_{(k)}$	Zdarzenia obserwowane dla $t_{(k)}$	Zdarzenia oczekiwane dla $t_{(k)}$
1	$n_{1k}$	$d_{1k}$	$e_{1k} = d_k (n_{1k} / n_k)$
2	$n_{2k}$	$d_{2k}$	$e_{2k} = d_k (n_{2k} / n_k)$
Razem	$n_k$	$d_k$	$e_k = e_{1k} + e_{2k} = d_k$

- ◆ Intuicja: przy założeniu hipotezy zerowej, zdarzenia powinny być rozłożone proporcjonalnie do zbioru ryzyka.
- ◆ Czyli  
oczekiwane = zaobserwowane \* (zb. ryzyka dla grupy / sumaryczny)

# Dwupróbkowy test logrank (5)

Event time	Total risk set	Total events	Group 1			Group 2			Variances
			Risk set	Observed events	Expected events	Risk set	Observed events	Expected events	
$t_{(1)}$	$n_1$	$d_1$	$n_{11}$	$d_{11}$	$e_{11}=d_1(n_{11}/n_1)$	$n_{21}$	$d_{21}$	$e_{21}=d_1(n_{12}/n_1)$	$V_{11}, V_{21}$
$t_{(2)}$	$n_2$	$d_2$	$n_{12}$	$d_{12}$	$e_{12}=d_2(n_{12}/n_2)$	$n_{22}$	$d_{22}$	$e_{22}=d_2(n_{22}/n_2)$	$V_{12}, V_{22}$
...	...	...	...	...	...	...	...	...	...
$t_{(K)}$	$n_K$	$d_K$	$n_{1K}$	$d_{1K}$	$e_{1K}=d_K(n_{1K}/n_K)$	$n_{2K}$	$d_{2K}$	$e_{2K}=d_K(n_{2K}/n_K)$	$V_{1K}, V_{2K}$
$\sum_k$				$d_{1+}$	$e_{1+}$		$d_{2+}$	$e_{2+}$	

- UWAGA:  $\sum_g d_{g+} = \sum_g e_{g+}$  (suma zaobserwowanych = suma oczekiwanych)

- ◆ Statystyka testowa oparta na różnicy (zaobserwowane-oczekiwane) w jednej z grup:

$$U_L = \sum_k (d_{1k} - e_{1k}) = \left( \sum_k d_{1k} \right) - \left( \sum_k e_{1k} \right) = d_{1+} - e_{1+}$$

# Dwupróbkowy test logrank (6)

Event time	Total risk set	Total events	Group 1			Group 2			Variances
			Risk set	Observed events	Expected events	Risk set	Observed events	Expected events	
$t_{(1)}$	$n_1$	$d_1$	$n_{11}$	$d_{11}$	$e_{11}=d_1(n_{11}/n_1)$	$n_{21}$	$d_{21}$	$e_{21}=d_1(n_{12}/n_1)$	$V_{11}, V_{21}$
$t_{(2)}$	$n_2$	$d_2$	$n_{12}$	$d_{12}$	$e_{12}=d_2(n_{12}/n_2)$	$n_{22}$	$d_{22}$	$e_{22}=d_2(n_{22}/n_2)$	$V_{12}, V_{22}$
...	...	...	...	...	...	...	...	...	...
$t_{(K)}$	$n_K$	$d_K$	$n_{1K}$	$d_{1K}$	$e_{1K}=d_K(n_{1K}/n_K)$	$n_{2K}$	$d_{2K}$	$e_{2K}=d_K(n_{2K}/n_K)$	$V_{1K}, V_{2K}$
$\Sigma_k$				$d_{1+}$	$e_{1+}$		$d_{2+}$	$e_{2+}$	

- ◆ Gdyby tabele były niezależne, wariancja  $U_L$  byłaby równa

$$V_L = \sum_k V_{1k}$$

- ◆ Nie są. Ale można pokazać, że różnice  $d_{1k}-e_{1k}$  są nieskorelowane, a  $V_L$  jest oszacowaniem ich wariancji.

# Dwupróbkowy test logrank: rozkład p-stwa statystyki testowej

- ◆ Przy założeniu hipotezy zerowej,

$$X_L^2 = U_L^2 / V_L$$

ma rozkład  $\chi^2$  z 1 stopniem swobody

- ◆  $P(X_L^2 \geq 3.84) = 0.05$

# Choroba lokomocyjna: test logrank

- ◆ Poziom istotności  $\alpha = 5\%$
- ◆ Zaobserwowana liczba zdarzeń:  $d_{1+}=5$ ,  $d_{2+}=14$   
Oczekiwana:  $e_{1+}=8.86$ ,  $e_{2+}=10.14$
- ◆ Statystyka testowa:  $X_L^2 = (5 - 8.86)^2 / 4.65 = 3.21$
- ◆  $p = P(X_L^2 \geq 3.21) = 0.073 > \alpha$ 
  - $X_L^2 < 3.84$
- ◆ Wynik nie jest istotny statystycznie – nie możemy odrzucić hipotezy zerowej

# Test logrank: prostsze obliczenia

◆ Zachodzi

$$X^2_L \approx \frac{(d_{1+} - e_{1+})^2}{e_{1+}} + \frac{(d_{2+} - e_{2+})^2}{e_{2+}}$$

- ◆ Dla badania choroby lokomocyjnej otrzymujemy

$$X_L^2 \approx (5 - 8.86)^2 / 8.86 + (14 - 10.14)^2 / 10.14 = 3.15$$

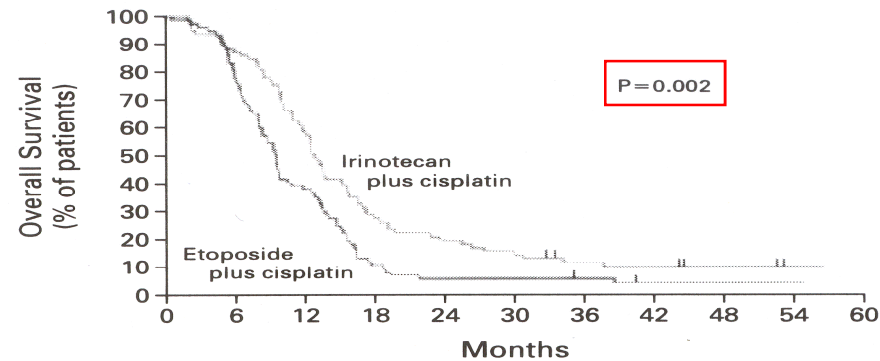
◆  $p = P(X_L^2 \geq 3.15) = 0.076 > \alpha$

•  $X_L^2 < 3.84$

- ◆ Nie możemy odrzucić hipotezy zerowej

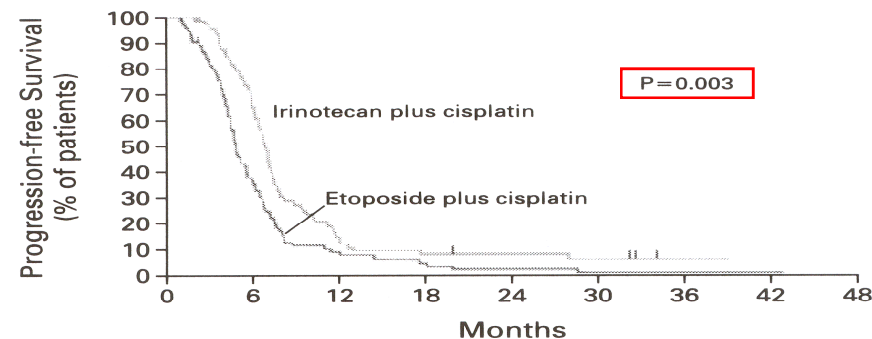
# Test logrank: próba kliniczna SCLC

Noda *et al.*,  
*NEJM* (2002)



No. AT RISK							
Irinotecan plus cisplatin	77	67	45	21	15	11	7
Etoposide plus cisplatin	77	60	29	8	4	4	3

**Figure 1.** Overall Survival of Patients with Extensive Small-Cell Lung Cancer Who Were Assigned to Treatment with Irinotecan plus Cisplatin or Etoposide plus Cisplatin. The tick marks indicate patients whose data were censored.



No. AT RISK				
Irinotecan plus cisplatin	77	47	9	9
Etoposide plus cisplatin	77	27	6	2

**Figure 2.** Progression-free Survival of Patients with Extensive Small-Cell Lung Cancer Who Were Assigned to Treatment with Irinotecan plus Cisplatin or Etoposide plus Cisplatin. The tick marks indicate patients whose data were censored.



# Alternatywy dla testu logrank

- ◆ Test logrank jest „wrażliwy” na „późne” różnice w krzywych przeżycia (dla większych wartości czasu).
- ◆ „Wczesne” różnice mogą być interesujące.
- ◆ Test *Wilcoxon*-*Gehan* jest „wrażliwy” na „wczesne” różnice.
- ◆ Uwaga: wybór testu powinien być dokonany przed oceną danych!

# Dwupróbkowy test Wilcoxona-Gehana

Event time	Total risk set	Total events	Group 1			Group 2			Variances
			Risk set	Observed events	Expected events	Risk set	Observed events	Expected events	
$t_{(1)}$	$n_1$	$d_1$	$n_{11}$	$d_{11}$	$e_{11}=d_1(n_{11}/n_1)$	$n_{21}$	$d_{21}$	$e_{21}=d_1(n_{12}/n_1)$	$V_{11}, V_{21}$
$t_{(2)}$	$n_2$	$d_2$	$n_{12}$	$d_{12}$	$e_{12}=d_2(n_{12}/n_2)$	$n_{22}$	$d_{22}$	$e_{22}=d_2(n_{22}/n_2)$	$V_{12}, V_{22}$
...	...	...	...	...	...	...	...	...	...
$t_{(K)}$	$n_K$	$d_K$	$n_{1K}$	$d_{1K}$	$e_{1K}=d_K(n_{1K}/n_K)$	$n_{2K}$	$d_{2K}$	$e_{2K}=d_K(n_{2K}/n_K)$	$V_{1K}, V_{2K}$
$\Sigma_k$				$d_{1+}$	$e_{1+}$		$d_{2+}$	$e_{2+}$	

- Statystyka testowa oparta na ważonych różnicach (zaobserwowane-oczekiwane) w jednej z grup:

$$U_W = \sum_k n_k (d_{1k} - e_{1k})$$

- większe wagi dla wcześniejszych czasów (większe zb. ryzyka)

- Wariancja statystyki:  $V_W = \sum_k n_k^2 V_{1k}$

# Dwupróbkowy test Wilcoxon-Gehana: rozkład p-stwa statystyki testowej

- ◆ Przy założeniu hipotezy zerowej,

$$X_W^2 = U_W^2 / V_W$$

ma rozkład  $\chi^2$  z 1 stopniem swobody

- ◆  $P(X_W^2 \geq 3.84) = 0.05$

# Alternatywne rodziny testów (1)

- ◆ Rozpatrujemy przypadek dwóch grup.
- ◆ Statystyka testowa oparta na ważonych różnicach (zaobserwowane-oczekiwane) w jednej z grup:

$$U = \sum_k \omega_k (d_{1k} - e_{1k})$$

- ◆ Różne wagi:
  - $\omega_k = 1 \rightarrow \text{logrank}$
  - $\omega_k = n_k \rightarrow \text{W-G}$
  - $\omega_k = (n_k)^{1/2} \rightarrow \text{Tarone-Ware}$
  - $\omega_k = S(t_{(k)}) \rightarrow \text{Peto-Peto-Prentice}$
  - $\omega_k = \{S(t_{(k)})\}^p \{1-S(t_{(k)})\}^q \rightarrow \text{Harrington-Fleming } (p, q \geq 0)$

# Alternatywne rodziny testów (2)

- ◆ Wagi „uwrażliwiają” na różnice w pewnych przedziałach czasu
  - logrank → „późne” różnice
  - W-G, T-W, P-P-P → „wczesne” różnice
- ◆ W-G i T-W zależą od cenzurowania
  - problem gdy wiele obserwacji cenzurowanych
- ◆ H-F bardzo „elastyczne”, ale jak wybrać  $p$  i  $q$  ?

# Test logrank dla więcej niż dwóch grup (1)

- ◆ Dla każdego czasu zdarzenia  $t_{(k)}$ , konstruujemy tabelę

Group	Risk set (alive) at $t_{(k)}$	Observed events at $t_{(k)}$	Expected events at $t_{(k)}$
1	$n_{1k}$	$d_{1k}$	$e_{1k} = d_k (n_{1k} / n_k)$
2	$n_{2k}$	$d_{2k}$	$e_{2k} = d_k (n_{2k} / n_k)$
...	...	...	...
G	$n_{Gk}$	$d_{Gk}$	$e_{Gk} = d_k (n_{Gk} / n_k)$
Total	$n_k$	$d_k$	$e_k = d_k$

- ◆ Rozkład warunkowy  $d_{1k}, \dots, d_{Gk}$ , przy danym  $d_k$ , jest rozkładem hipergeometrycznym:

$$\binom{n_k}{d_k}^{-1} \prod_{j=1}^G \binom{n_{jk}}{d_{jk}}$$

# Test logrank dla więcej niż dwóch grup (2)

- ◆ Rozważania dla dwóch grup uogólniają się:
  - Wektor różnic  $\mathbf{U}_L = (d_{1+} - e_{1+}, \dots, d_{G-1,+} - e_{G-1,+})'$
  - Macierz wariancji-kowariancji  $\mathbf{V}_L = \sum_k \mathbf{V}_k$  z rozkładu hipergeometrycznego
- ◆ Statystyka testowa:  $X_L^2 = \{\mathbf{U}_L\}^T \mathbf{V}_L^{-1} \mathbf{U}_L$ 
  - oparta na różnicach (zaobserwowane-oczekiwane) w  $G-1$  grupach.
- ◆ Przy założeniu hipotezy zerowej,  $X_L^2$  ma rozkład  $\chi^2$  z  $G-1$  stopniami swobody.

# Test logrank dla więcej niż dwóch grup: prostsze obliczenia

Group	Risk set (alive) at $t_{(k)}$	Observed events at $t_{(k)}$	Expected events at $t_{(k)}$
1	$n_{1k}$	$d_{1k}$	$e_{1k} = d_k (n_{1k} / n_k)$
2	$n_{2k}$	$d_{2k}$	$e_{2k} = d_k (n_{2k} / n_k)$
...	...	...	...
G	$n_{Gk}$	$d_{Gk}$	$e_{Gk} = d_k (n_{Gk} / n_k)$
Total	$n_k$	$d_k$	$e_k = d_k$

- ◆ Zachodzi

$$X_L^2 \approx \frac{(d_{1+} - e_{1+})^2}{e_{1+}} + \frac{(d_{2+} - e_{2+})^2}{e_{2+}} + \dots + \frac{(d_{G+} - e_{G+})^2}{e_{G+}}$$
- ◆ Przy założeniu hipotezy zerowej,  $X_L^2$  ma rozkład  $\chi^2$  z  $G-1$  stopniami swobody.



# Alternatywne rodziny testów dla więcej niż dwóch grup

- ◆ Podobnie jak w teście dwupróbkowym, używane są odpowiednio ważone różnice (zaobserwowane-oczekiwane):

- Wektor różnic  $\mathbf{U}=(U_1, \dots, U_G)'$ , gdzie  $U_g = \sum_k \omega_k (d_{gk} - e_{gk})$
- Macierz wariancji-kowariancji  $\mathbf{V}=\sum_k \omega_k^2 \mathbf{V}_k$  z rozkładu hipergeometrycznego

- ◆ Statystyka testowa:  $X^2 = \{\mathbf{U}\}^T \mathbf{V}^{-1} \mathbf{U}$

- ◆ Przy założeniu hipotezy zerowej,  $X^2$  ma rozkład  $\chi^2$  z  $G-1$  stopniami swobody.

# Test logrank dla uporządkowanych grup

- ◆ Dotyczy porównania krzywych przeżycia dla trzech lub więcej **uporządkowanych grup**:

- np. stopień zaawansowania guza (T0,T1,T2), grupy wieku,...

- ◆ **Test logrank dla trendu**

- Hipoteza zerowa:  $S_1(t)=S_2(t)=S_3(t)=\dots=S_G(t)$

- Hipoteza alternatywna:  $S_1(t)<S_2(t)<S_3(t)<\dots<S_G(t)$

- ◆ Dla porządkowej hipotezy alternatywnej testu dla trendu ma większą moc statystyczną niż „zwykły” test logrank

- Wyższe p-stwo odrzucenia hipotezy zerowej, gdy jest fałszywa

# Test logrank dla trendu

- ♦ Używamy wag oddających porządek grup:

- Np. 1, 2, 3, albo -1, 0, 1 dla trzech grup

- ♦ Ważone różnice (obs. – ocz.):  $U_{LT} = \sum_{g=1}^G w_g (d_{g+} - e_{g+})$

- ♦ Wariancja:  $V_{LT} = \sum_{g=1}^G w_g^2 e_{g+} - \frac{\left( \sum_{g=1}^G w_g e_{g+} \right)^2}{\sum_{g=1}^G e_{g+}}$

- ♦ Przy założeniu hipotezy zerowej,

$$X_{LT}^2 = U_{LT}^2 / V_{LT}$$

ma rozkład  $\chi^2$  z 1 stopniem swobody.

# Test logrank dla trendu: przykład (1)

- ◆ Logrank:  $p=0.06$
- ◆ P(przeżycie) maleje ze wzrostem liczby zajętych węzłów
- ◆ Stosujemy test dla trendu
- ◆ (Uwaga: decyzja powinna być podjęta przed oceną danych)

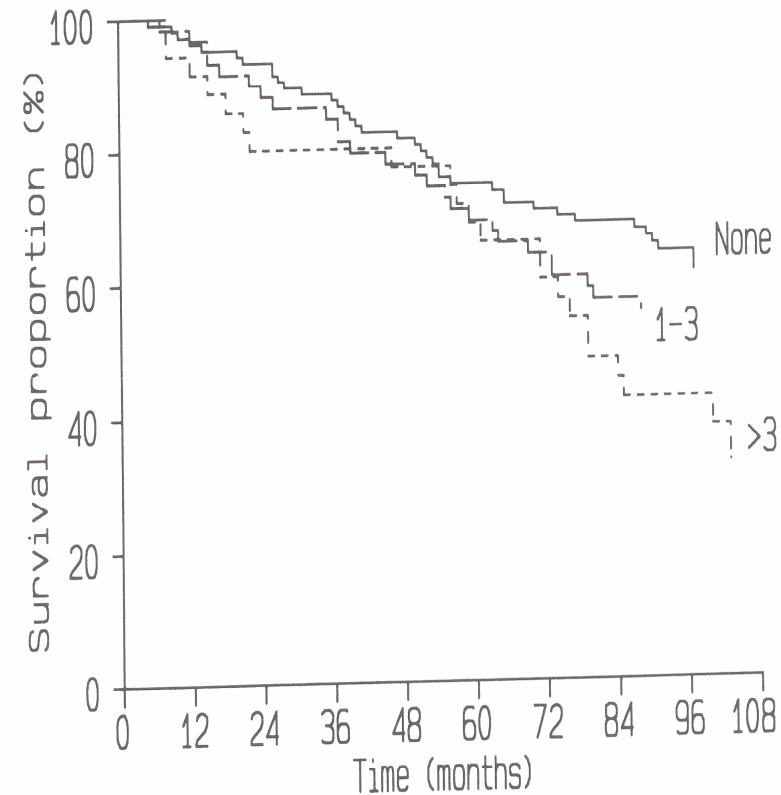


Figure 13.6 Kaplan-Meier curves for patients with breast cancer with none ( $n = 102$ ), 1-3 ( $n = 58$ ), or more than 3 ( $n = 35$ ) positive nodes (data from Barnes *et al.*, 1988).

# Test logrank dla trendu: przykład (2)

Positive nodes	# of women	$d_{g+}$	$e_{g+}$	$d_{g+} - e_{g+}$
none	102	38	46.41	-8.41
few (1-3)	58	26	25.21	0.79
many (>3)	35	22	14.38	7.62
Total	195	86	86	0

◆ Przyjmijmy wagi -1, 0, 1

Positive nodes	$w_g(d_{g+} - e_{g+})$	$w_g e_{g+}$	$w_g^2 e_{g+}$
none	8.41	-46.41	46.41
few (1-3)	0	0	0
many (>3)	7.62	14.38	14.38
Total	16.03	-32.03	60.79

$$V_{LT} = \sum_{g=1}^G w_g^2 e_{g+} - \frac{\left( \sum_{g=1}^G w_g e_{g+} \right)^2}{\sum_{g=1}^G e_{g+}}$$

- ◆ Otrzymujemy  $U_{LT} = 16.03$  oraz  $V_{LT} = 60.79 - (-32.03)^2 / 86 = 48.86$
- ◆  $X_{LT}^2 = U_{LT}^2 / V_{LT} = 16.03^2 / 48.86 = 5.26$
- ◆  $p = P(X_{LT}^2 \geq 5.26) = 0.02 < 0.05$ 
  - $X_{LT}^2 > 3.84$
- ◆ Wynik testu jest istotny statystycznie.

# Alternatywne rodziny testów dla trendu

- ◆ Podobna idea jak dla logrank: używamy wag  $w_g$  oddających porządek grup
  - Wektor różnic  $\mathbf{U}=(U_1, \dots, U_G)'$ , gdzie  $U_g = \sum \omega_k (d_{gk} - e_{gk})$
  - Ważona suma  $U_T = \mathbf{w}' \mathbf{U}$ , gdzie  $\mathbf{w}=(w_1, \dots, w_G)'$  – wektor wag
  - Wariancja  $V_T = \mathbf{w}' (\sum_k \omega_k^2 \mathbf{V}_k) \mathbf{w} = \mathbf{w}' \mathbf{V} \mathbf{w}$
- ◆ Statystyka testowa:  $X_T^2 = U_T^2 / V_T$
- ◆ Przy założeniu hipotezy zerowej,  $X_T^2$  ma rozkład  $\chi^2$  z 1 stopniem swobody.

# Warstwowy test logrank (1)

- ◆ Pozwala na uwzględnienie wpływu zmiennych towarzyszących:
  - oczekiwane liczby zdarzeń liczone są osobno w każdej z warstw...
  - ... a następnie sumowane i porównywane z zaobserwowanymi, jak w „zwykłym” teście logrank.

## Warstwowy test logrank (2)

	Zdarzenia, grupa 1		...	Zdarzenia, grupa G	
Warstwa	obs	ocz	...	obs.	ocz.
1	$d_{11+}$	$e_{11+}$	...	$d_{1G+}$	$e_{1G+}$
...	...	...	...	...	...
S	$d_{S1+}$	$e_{S1+}$	...	$d_{SG+}$	$e_{SG+}$
Razem	$d_{+1+}$	$e_{+1+}$	...	$d_{+G+}$	$e_{+G+}$

♦ Statystyka testowa w warstwie s:  $\{\mathbf{U}_L^{(s)}\}^T \{\mathbf{V}_L^{(s)}\}^{-1} \mathbf{U}_L^{(s)}$

♦ Zdefiniujmy 
$$X_{SL}^2 = \left( \sum_{s=1}^S \mathbf{U}_L^{(s)} \right)^T \left( \sum_{s=1}^S \mathbf{V}_L^{(s)} \right)^{-1} \left( \sum_{s=1}^S \mathbf{U}_L^{(s)} \right)$$

♦ Przy założeniu hipotezy zerowej,  $X_{SL}^2$  ma rozkład  $\chi^2$  z  $G-1$  stopniami swobody.



# Warstwowy test logrank: prostsze obliczenia

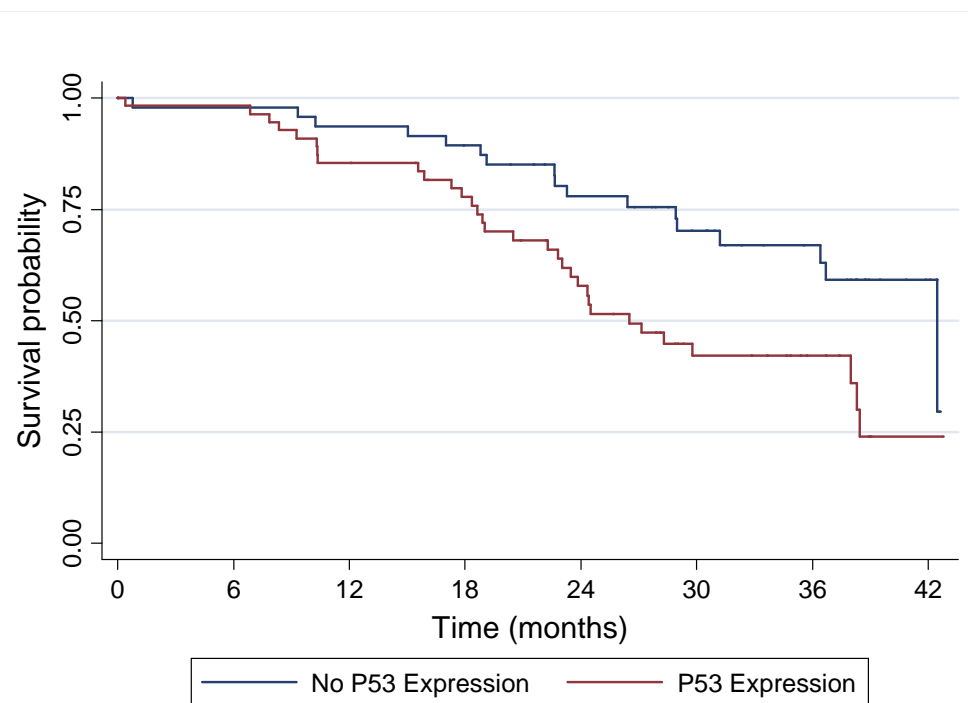
Warstwa	Zdarzenia, grupa 1		...	Zdarzenia, grupa G	
	obs	ocz		obs.	ocz.
1	$d_{11+}$	$e_{11+}$	...	$d_{1G+}$	$e_{1G+}$
...	...	...	...	...	...
S	$d_{S1+}$	$e_{S1+}$	...	$d_{SG+}$	$e_{SG+}$
Razem	$d_{+1+}$	$e_{+1+}$	...	$d_{+G+}$	$e_{+G+}$

$$X_{SL}^2 \approx \frac{\left\{ \sum_{s=1}^S (d_{s1+} - e_{s1+}) \right\}^2}{\sum_{s=1}^S e_{s1+}} + \dots + \frac{\left\{ \sum_{s=1}^S (d_{sG+} - e_{sG+}) \right\}^2}{\sum_{s=1}^S e_{sG+}} = \frac{(d_{+1+} - e_{+1+})^2}{e_{+1+}} + \dots + \frac{(d_{+G+} - e_{+G+})^2}{e_{+G+}}$$

- ◆ Przy założeniu hipotezy zerowej,  $X_{SL}^2$  ma rozkład  $\chi^2$  z  $G-1$  stopniami swobody.

# Badanie NSCLC: ekspresja białka P53

- ♦ Laudański *et al.*, *Eur Respir J* (2001)
- ♦ 102 chorych leczonych chirurgicznie
- ♦ Ekspresja białka P53 z biopsji guza
- ♦ Wpływ na czas przeżycia
- ♦ Logrank:  $p=0.008 \rightarrow$

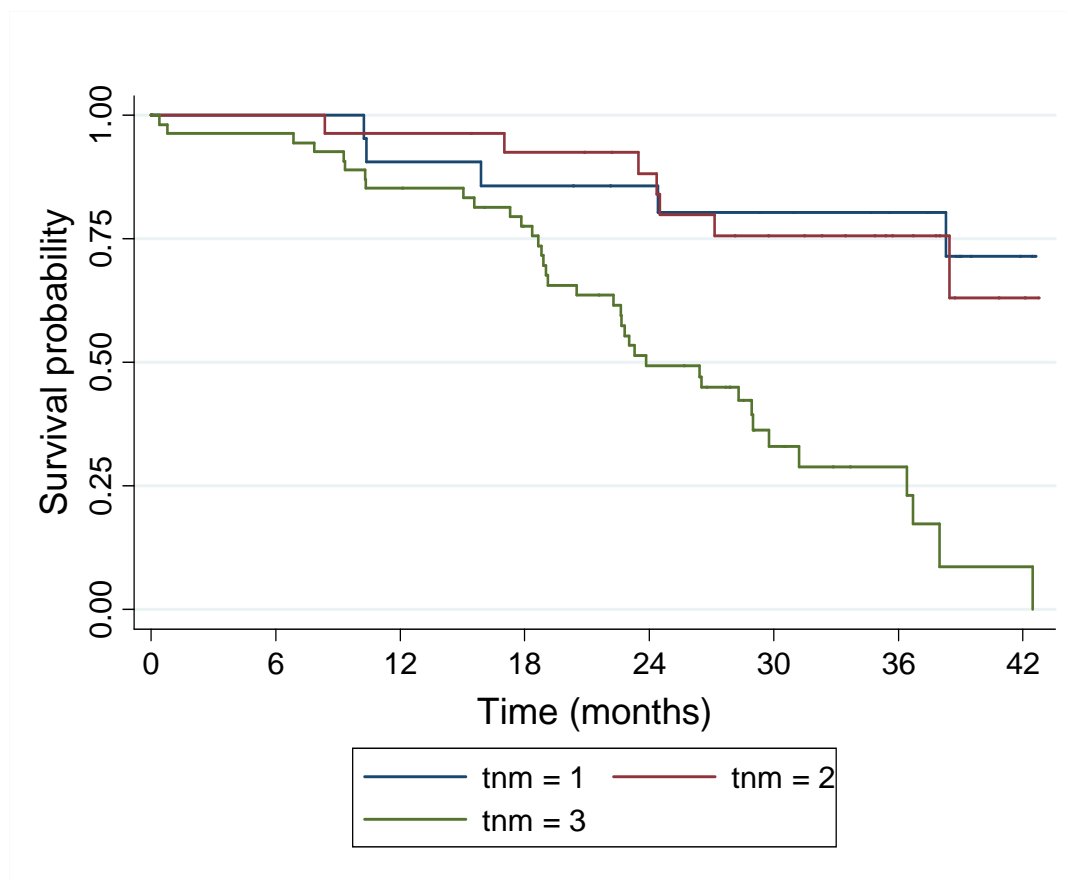


# Badanie NSCLC: TNM

♦ TNM, logrank dla trendu:  
 $p < 0.001$

♦ TNM ma wpływ na czas przeżycia

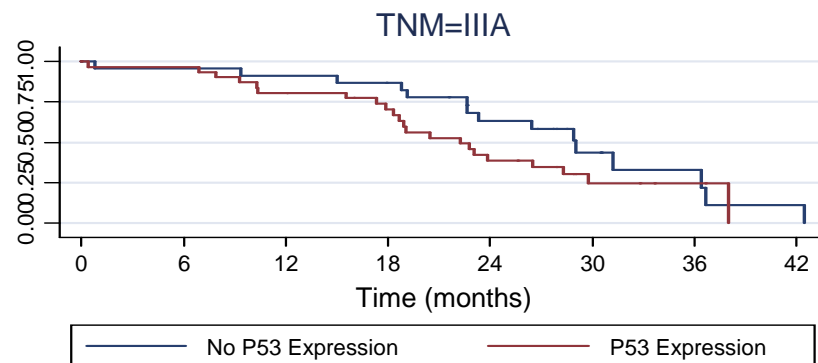
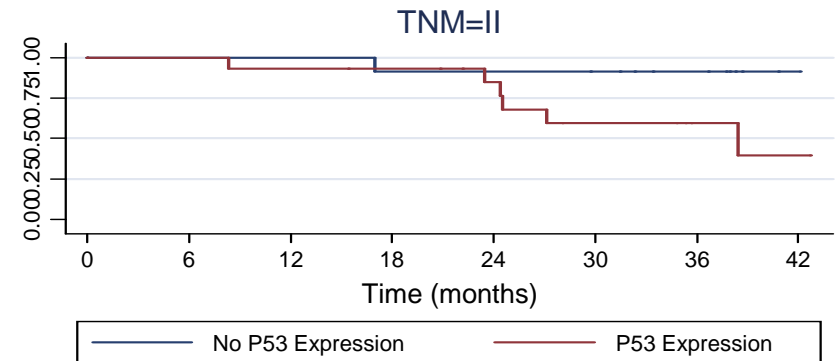
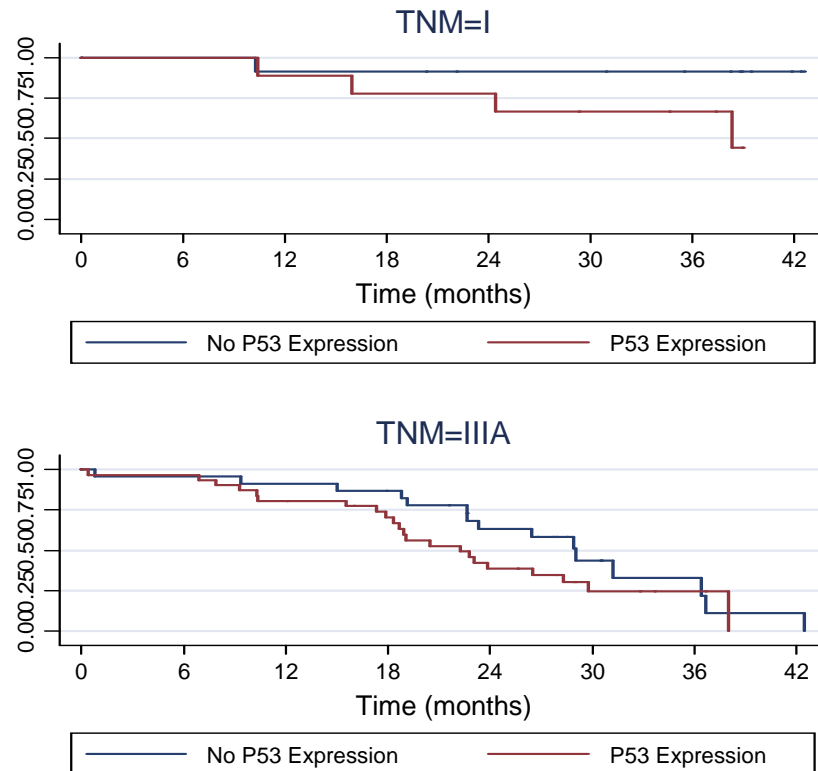
TNM	expression+
I	42.9%
II	55.6%
IIIA	57.4%



♦ Związek między TNM a dodatnią ekspresją białka P53

♦ Wpływ ekspresji na przeżycie może wynikać z efektu TNM

# Warstwowy test logrank: NSCLC (1)



♦ Warstwowy: TNM

TNM=1	Events	Events
expr	observed	expected
0	1	2.93
1	4	2.07
Total	5	5.00

TNM=2	Events	Events
expr	observed	expected
0	1	3.58
1	6	3.42
Total	7	7.00

TNM=3	Events	Events
expr	observed	expected
0	15	18.97
1	22	18.03
Total	37	37.00

# Warstwowy test logrank: NSCLC (2)

Warstwa	Zdarzenia, grupa 1		Zdarzenia, grupa 2	
	Obs	Ocz	Obs	Ocz
1	1	2.93	4	2.07
2	1	3.58	6	3.42
3	15	18.97	22	18.03
Razem	17	25.48	32	23.52

## ♦ Dwie grupy

♦  $X_{SL}^2 \approx (17-25.48)^2/25.48 + (32-23.52)^2/23.52 = 5.88$

•  $X_{SL}^2$  ma rozkład  $\chi^2$  z 1 stopniem swobody

♦  $p = P(X_{SL}^2 \geq 5.88) = 0.015 < 0.05$

•  $X_{SL}^2 > 3.84$

• Dla wzoru “skomplikowanego” :  $X_{SL}^2 = 6.05$ ;  $p=0.014$

♦ Efekt ekspresji P53 wciąż istotny statystycznie, ale mniejszy poziom krytyczny testu.

# Warstwowanie

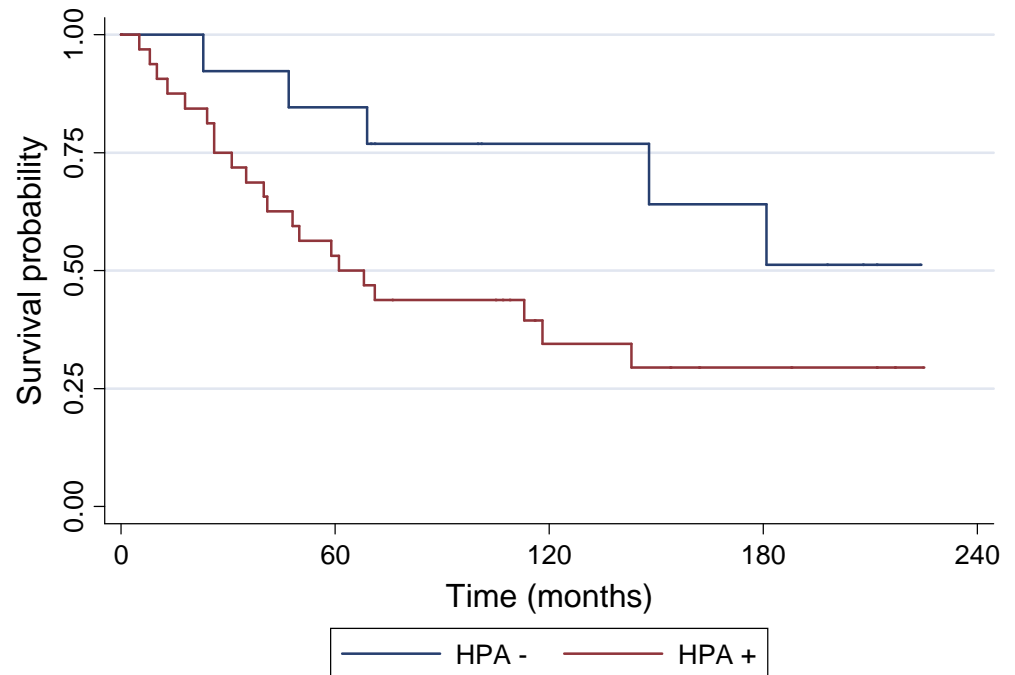
- ◆ Może być użyte dla więcej niż jednej zmiennej towarzyszącej
  - ale im więcej warstw, tym mniejsza liczba obserwacji w warstwie
- ◆ Może być stosowane dla alternatywnych testów, np. Wilcoxon-Gehana etc. Ogólna idea jest podobna: liczymy statystyki testowe osobno w warstwach, a następnie je łączymy.

# Rak piersi i HPA

- ◆ Leathem & Brooks, *Lancet* (1987)
- ◆ HPA – marker używany do znaczenia komórek raka piersi z przerzutami do węzłów chłonnych
- ◆ Badanie retrospektywne: 45 kobiet po leczeniu chirurgicznym; próbki guza znaczone HPA i oceniane pod kątem istnienia przerzutów do węzłów chłonnych
- ◆ Czy wynik oceny HPA jest czynnikiem rokowniczym (przerzuty do węzłów = gorsze rokowanie)?

# Test Wilcoxon-Gehana i test logrank: rak piersi i HPA

- ◆ Logrank:  $p=0.061$
- ◆ W-G:  $p=0.041$
- ◆ Nieco różne wyniki



- ◆ (Tylko przykład!: test powinien być wybrany przed oceną danych)



# Test Wilcoxona-Gehana i logrank (1)

- ◆ Rak piersi i HPA: logrank:  $p=0.061$ ; W-G:  $p=0.041$
- ◆ Czy nie można użyć *obu* testów?
- ◆ Problem:
  - dwa testy → poprawka Bonferroniego ( $\alpha = 0.05/2$ )
  - testy skorelowane → poprawka Bonferroniego jest konserwatywna

# Test Wilcoxona-Gehana i logrank (2)

- ◆ Tarone (1981): max(logrank, W-G)
- ◆ Statystyki testowe dla dwupróbkowych testów:

logrank:  $X_L^2 = U_L^2 V_L^{-1}$

Wilcoxona-Gehana:  $X_W^2 = U_W^2 V_W^{-1}$

Tarone'a-Ware'a:  $X_{TW}^2 = U_{TW}^2 V_{TW}^{-1}$

- ◆  $P(\max(X_L^2, X_W^2) \leq x) = P\{ (-x \leq X_L \leq x) \cap (-x \leq X_W \leq x) \}$
- ◆ Rozkład  $(X_L, X_W)$  przybliżony dwuwymiarowym normalnym ze średnimi 0, wariancjami 1 i  $\text{Corr}(X_L, X_W)$ .
  - Oszacowanie  $\text{Corr}(X_L, X_W)$ :  $V_{TW} / (V_L V_W)^{1/2}$

# Test Wilcoxon-Gehana i logrank (3)

## ♦ Rak piersi i HPA:

- logrank:  $X_L^2 = (4.565)^2/5.929 = 3.515$ ,  $p=0.061$
- W-G :  $X_W^2 = (-159)^2/6048.135 = 4.180$ ,  $p=0.041$
- T-W :  $X_{TW}^2 = (-26.922)^2/178.865 = 4.052$ ,  $p=0.044$

♦  $\max(X_L^2, X_W^2) = 4.180$

♦  $\text{Corr}(X_L, X_W) = 178.865/(5.929 \cdot 6048.135)^{1/2} = 0.944$

♦  $P(\max(X_L^2, X_W^2) \geq 4.180) = p = 0.054$

# Ograniczenia analizy jednej zmiennej

- ◆ Porównanie krzywych przeżycia odpowiadających poziomom jednej zmiennej może być obciążone (zakłócone) wpływem innej zmiennej.
  - Przykład: efekt ekspresji P53 i TNM dla NSCLC
- ◆ Warstwowy test logrank może być pomocny, ale jedynie dla ograniczonej liczby warstw.
- ◆ Rozwiązanie: użycie modeli