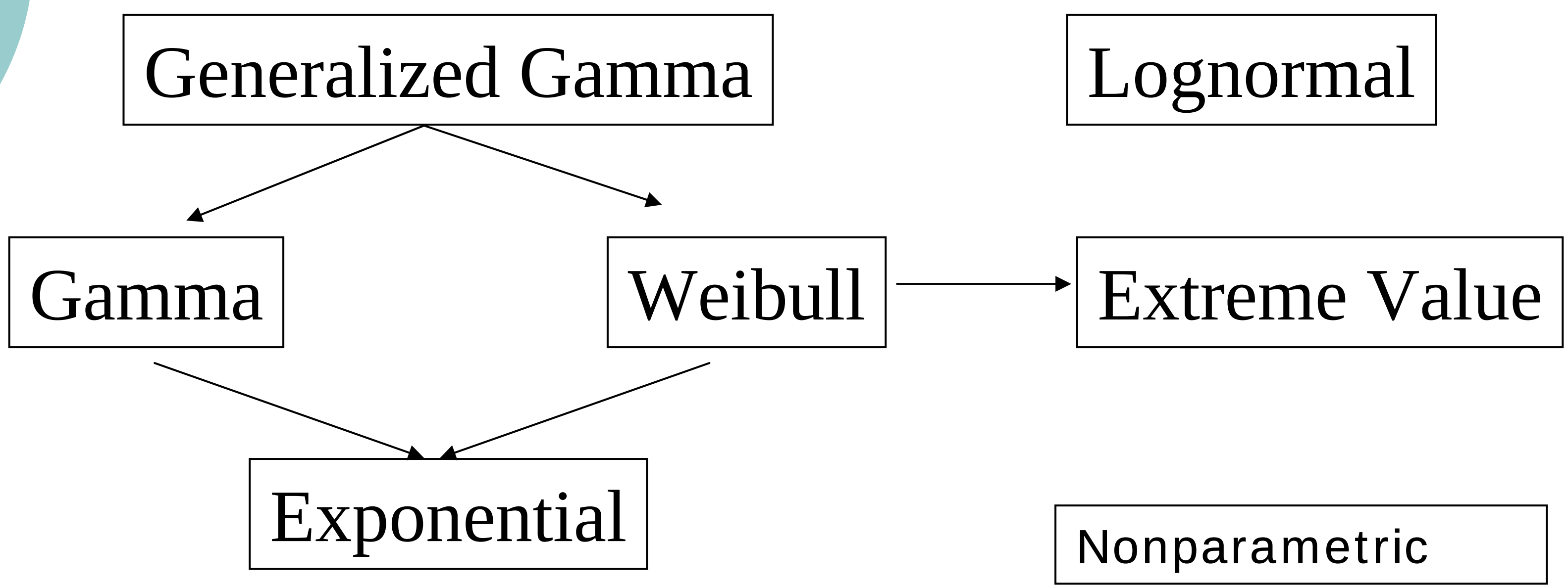
MULTIVARIATE SURVIVAL DISTRIBUTIONS

The Role of Statistical Modeling in Modern Research

Mark Carpenter
Director of Statistics
Department of Mathematics and Statistics

Univariate Survival Models are well developed



Lifetimes within a population are typically modeled with probability distributions possessing **positive supports**, i.e., $P(X>0)=1$.

Multivariate Survival Distributions

$$\begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_m \end{pmatrix} \longrightarrow \begin{matrix} T_1 & \text{(Event 1)} \\ T_2 & \text{(Event 2)} \\ \vdots & \vdots \\ T_m & \text{(Event m)} \end{matrix}$$

Multivariate Response — Univariate Responses

$$f(t_1, t_2 \cdots, t_m), \quad \vec{t} \in \Re^{m+}$$

$$f(t_1) = \iint \cdots \int f(y_2, y_3 \cdots, y_m) dy_2 dy_3 \cdots dy_m$$

Multivariate Survival Distributions

$$\begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_m \end{pmatrix} \longrightarrow \begin{matrix} T_1 & \text{(Event 1)} \\ T_2 & \text{(Event 2)} \\ \vdots & \vdots \\ T_m & \text{(Event m)} \end{matrix}$$

Multivariate Response   Univariate Responses

$$S(t_1, t_2, \ldots, t_m) = P(T_1 > t_1, T_2 > t_2, \ldots, T_m > t_m)$$

$$= \int_{t_1}^{\infty} \int_{t_2}^{\infty} \cdots \int_{t_m}^{\infty} f(y_1, y_2, \ldots, y_m) dy_1 dy_2 \ldots dy_m$$

Example Multivariate Situations

- **Internet Traffic:** Interarrival times correlated with service times (bivariate)
- **Tumor Incidence:** Time to first tumor correlated with time to second tumor (bivariate)
- **Competing Risks:** Time to organ failure in Diabetics (multivariate)

Gamma Random pdf

We say that $X \sim Ga(\lambda, \alpha, \mu)$ if

$$f_X(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}(x-\mu)^{\alpha-1} e^{-\lambda(x-\mu)^{\alpha}} I_{[\mu,\infty)}(x),$$

where $\mu \in \Re$ and $\lambda, \alpha \in \Re^{+}$.

Gamma Random pdf

We say that $X \sim Ga(\lambda, \alpha, \mu)$ if

$$f_X(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}(x-\mu)^{\alpha-1}e^{-\lambda(x-\mu)^{\alpha}} I_{[\mu,\infty)}(x),$$

where $\mu \in \Re$ and $\lambda, \alpha \in \Re^{+}$.

Shape

Scale

Location

Generating BVG with Linear Associations

- We start out with Pair-wise Associations

$$X_2 = aX_1 + Z, \quad a > 0$$

$$Ga(\lambda_1, \alpha_1, \mu_1)$$

$$Ga(\lambda_2, \alpha_2, \mu_2)$$

$$\text{If } a = \frac{\lambda_1}{\lambda_2} \text{ then } Z \sim Ga(\lambda_2, \alpha_2 - \alpha_1, \mu_2 - a\mu_1)$$

Generating BVG with Linear Associations

- We start out with Pair-wise Associations

$$X_2 = aX_1 + Z \Rightarrow Z = X_2 - aX_1$$

$$L_Z(s) = \frac{L_{X_2}(s)}{L_{X_1}(as)} = \left(\frac{\lambda_2}{\lambda_2 + s}\right)^{\alpha_2} \left(\frac{\lambda_1}{\lambda_1 + as}\right)^{-\alpha_1}$$

$$= \left(\frac{\lambda_2}{\lambda_2 + s}\right)^{\alpha_2} \left(\frac{\lambda_1}{\lambda_1 + (\lambda_1 / \lambda_2)s}\right)^{-\alpha_1} = \left(\frac{\lambda_2}{\lambda_2 + s}\right)^{\alpha_2 - \alpha_1}$$

Generating BVG with Linear Associations

- We start out with Pair-wise Associations

$$X_2 = aX_1 + Z, \quad a > 0$$

$$Z \text{ is a GAMMA } \textit{iff} \;\; \mathrm{a} = \frac{\lambda_1}{\lambda_2}$$

Generating BVG with Linear Associations

- We start out with Pair-wise Associations

$$X_2 = aX_1 + Z, \quad a > 0$$

$$Z \text{ is a GAMMA } \textit{iff} \text{ a} = \frac{\lambda_1}{\lambda_2}$$

Carpenter, Diawara and Han (2006)
Mathai and Moschopoulos (1991,1992)

$MVG(\alpha_i, \lambda_i, \mu_i, i = 1, \cdots, p)$

Let $X_0, Z_1, \ldots, Z_p$ be independent unobservable random variables. Let

$X_i \sim Ga(\lambda_i, \alpha_i, \mu_i), i = 1, \ldots, p$

and $X_i = a_i X_0 + Z_i$ then

$X' = (X_1, \ldots, X_p)$ is MVG.

BVG with Linear Associations

What if $a \neq \lambda_0 / \lambda_1$???

BVG with Linear Associations

What if $a \neq \lambda_0 / \lambda_1$???

We focus on a special case.

What if $a \neq \lambda_0 / \lambda_1$???

$$X_2 = aX_1 + Z, \quad a > 0, X_i \sim \exp(\lambda_i)$$

$$L_Z(s) = \frac{L_{X_2}(s)}{L_{X_1}(as)} = \left(\frac{\lambda_2}{\lambda_2 + s}\right)\left(\frac{\lambda_1}{\lambda_1 + as}\right)^{-1}$$

$$= p + (1 - p)L_{X_2}(s)$$

where $p = a\lambda_2 / \lambda_1$

Linearly Related Bivariate Weibull Distributions
Carpenter, Diawara and Han (2005)

- The Latent Variable Z associates X_2 to X_1.

$$X_2 = aX_1 + Z = \begin{cases} aX_1 & \text{if} \quad Z = 0 \\ aX_1 + Z & \text{if} \quad Z \neq 0 \end{cases}$$

- pdf of the Latent Variable Z

$$\delta(x) = 0, \quad \text{if} \quad x \neq 0,$$

$$\text{and} \quad \int_R \delta(t)\,dt = 1$$

$$f_Z(z) = p\delta(z) + (1-p) f_{X_2}(z) I(z > 0);$$

$$\text{where} \quad p = P(Z = 0) > 0.$$

- Joint MLE's of (λ1, λ2)

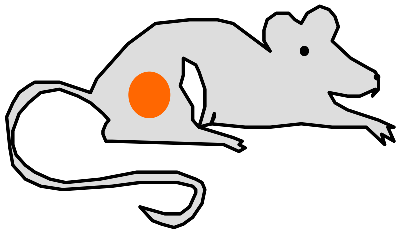
$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \dfrac{a}{\overline{x}_2} + \dfrac{n-k}{n\overline{x}_1} \\ \dfrac{1}{\overline{x}_2} \end{pmatrix}, where\ k = \sum_{i=1}^{n} I(x_2 - ax_1 = 0)$$

$$\text{var/cov} = \frac{1}{n} \begin{pmatrix} \lambda_1(\lambda_1 - a\lambda_2) + a^2\lambda_2^2 & a\lambda_2^2 \\ a\lambda_2^2 & \lambda_2^2 \end{pmatrix}$$

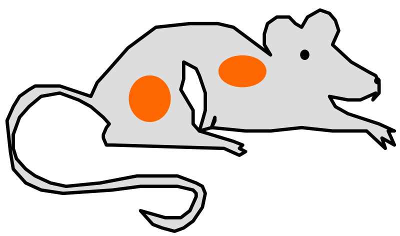
Simulation Study on Unit Exponential Margins

	MSE			MSE		
Corr			%improv			%improv
0.01	0.046228	0.046836	1.2981	8.63E-06	0.000009	4.1333
0.05	0.043563	0.045831	4.9486	0.000189	0.000197	4.2660
0.10	0.046951	0.049607	5.3541	0.000749	0.000813	7.8133
0.20	0.041918	0.047798	12.3018	0.002704	0.003054	11.4701
0.30	0.0364	0.048083	24.2976	0.004359	0.005422	19.6116
0.40	0.034673	0.045331	23.5115	0.006471	0.008153	20.6275
0.50	0.03859	0.052588	26.6182	0.007359	0.009298	20.8570
0.60	0.039866	0.051227	22.1778	0.007539	0.011078	31.9430
0.70	0.040424	0.051023	20.7730	0.007635	0.011208	31.8773
0.80	0.039993	0.045867	12.8066	0.005708	0.008814	35.2357
0.90	0.045294	0.051646	12.2991	0.00345	0.005859	41.1096
0.99	0.048218	0.048155	-0.1308	0.000356	0.000639	44.3100

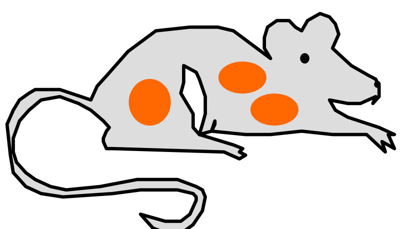
Chemoprevention Example



T_1 = time to first tumor



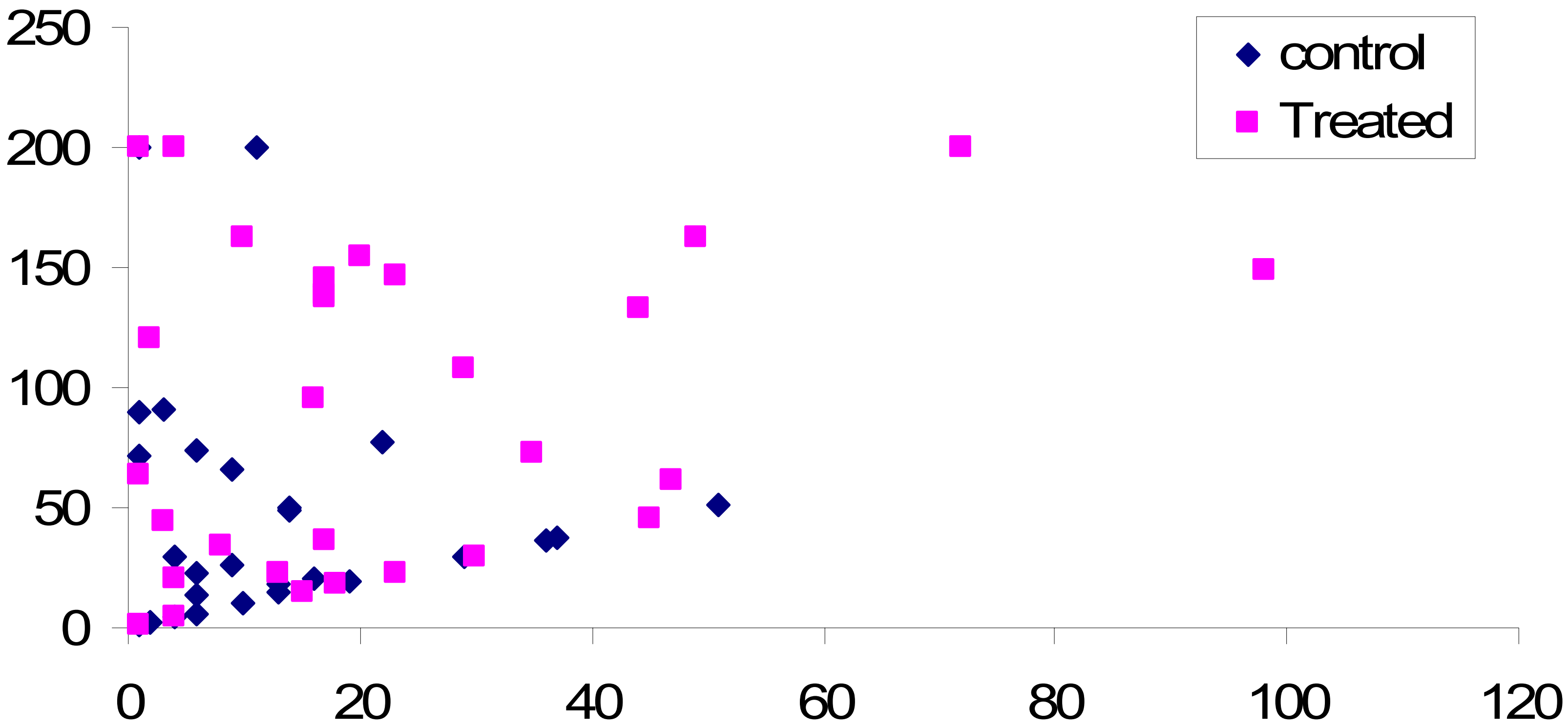
T_2 = time to second tumor



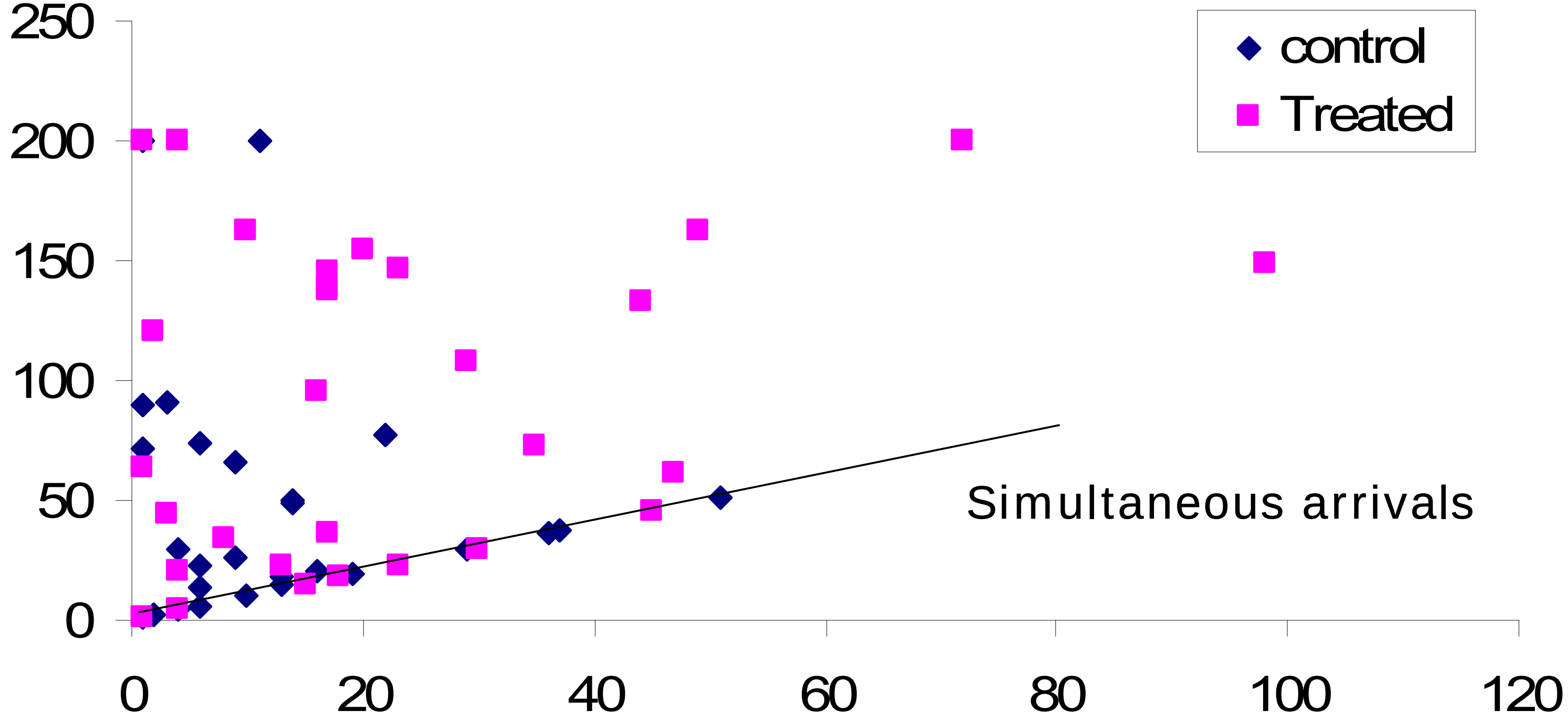
T_3 = time to third tumor

T_4 = time to death

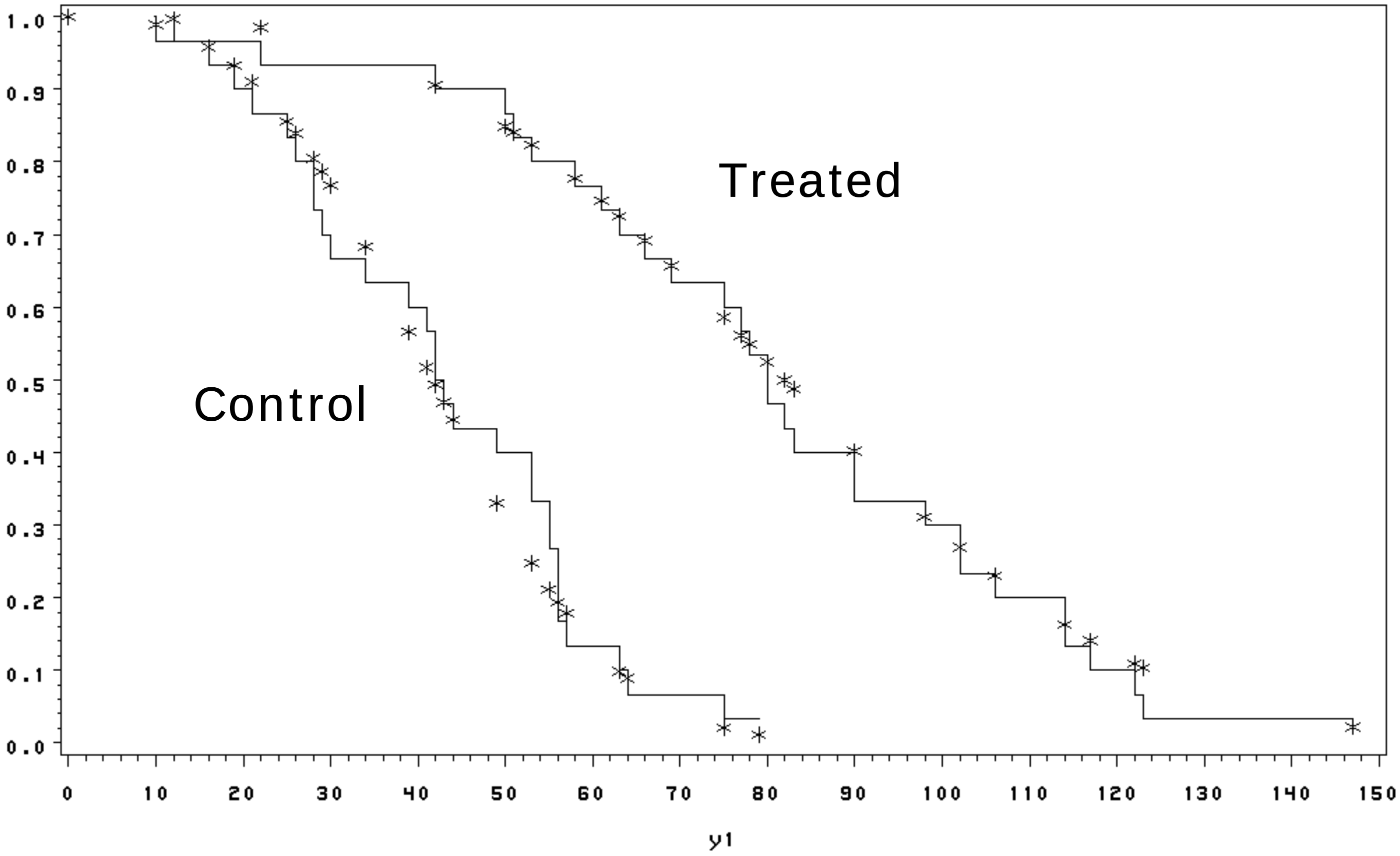
Scatter Plots of first/second tumor times



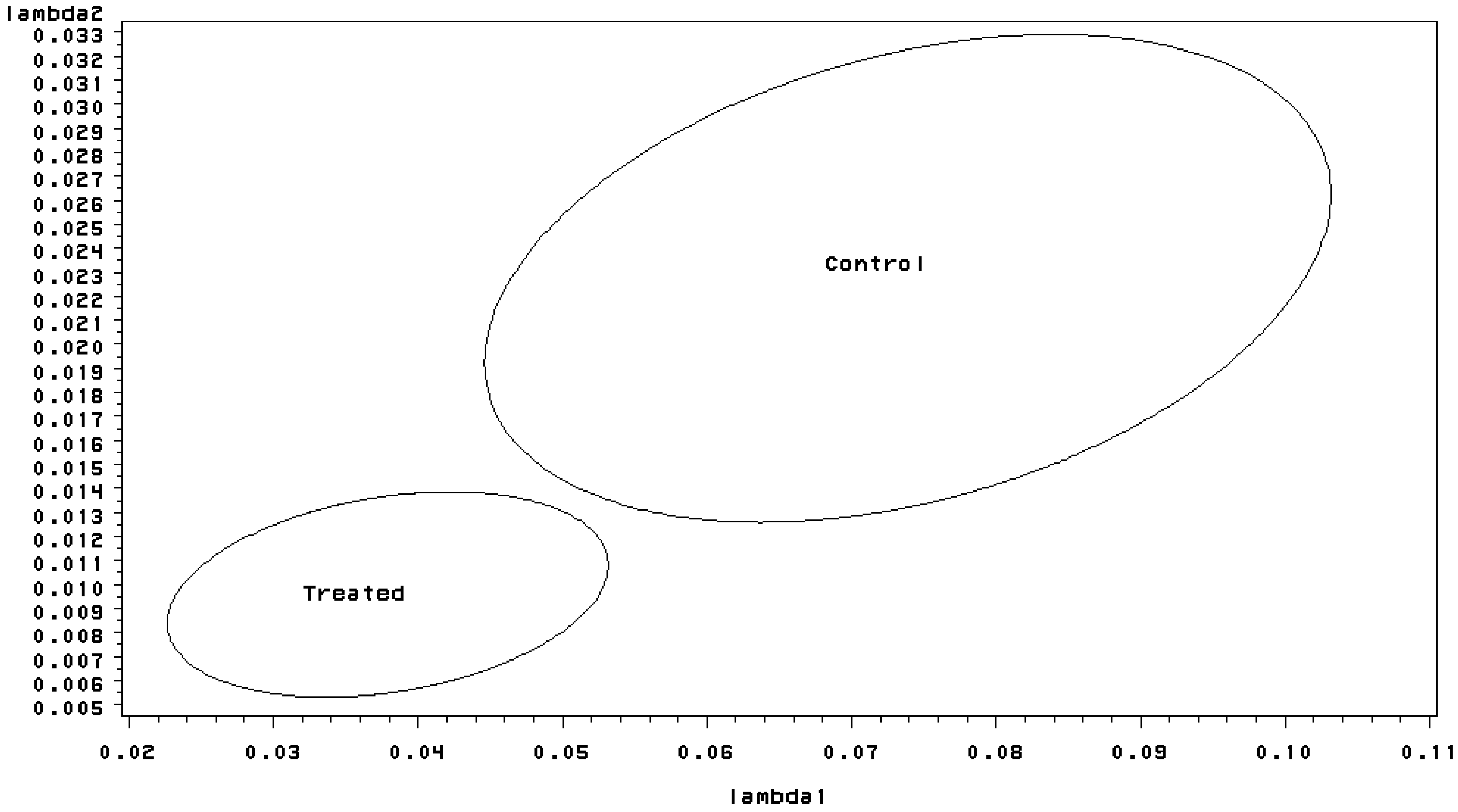
Scatter Plots of first/second tumor times



Time to tumor



Confidence Ellipsoids



$$\text{mean time to tumor} = 1/\lambda$$

Conclusions

- Ignoring correlation between events is inefficient and can lead to difficult to interpret results
- More general models are necessary to broaden the list of applications

Current Research

- Development of a generalized multivariate gamma, location-scale, family and parameter estimation (Norou Diawara)
- Development of a bivariate Weibull regression model and parameter estimation (Yi Han)
- Mixture models to solve the internet intruder clustering problem.