New Developments in Survival Analysis Using SAS® Software

Joseph C. Gardiner

Division of Biostatistics, Department of Epidemiology,

Michigan State University, East Lansing, MI

MISUG September 23, 2010, Ann Arbor, Michigan

INTRODUCTION

- **Survival Analysis– methods to analyze time-to-event data that are right, left and/or interval censored**

- **Extensions from single event to repeated (recurrent) events, and multiple events**

- **Methods for different types of analyses:**

  I. Nonparametric†
  II. Parametric†
  III. Semi-parametric
  IV. Bayesian

† Focus of this presentation

I. NONPARAMETRIC ANALYSIS

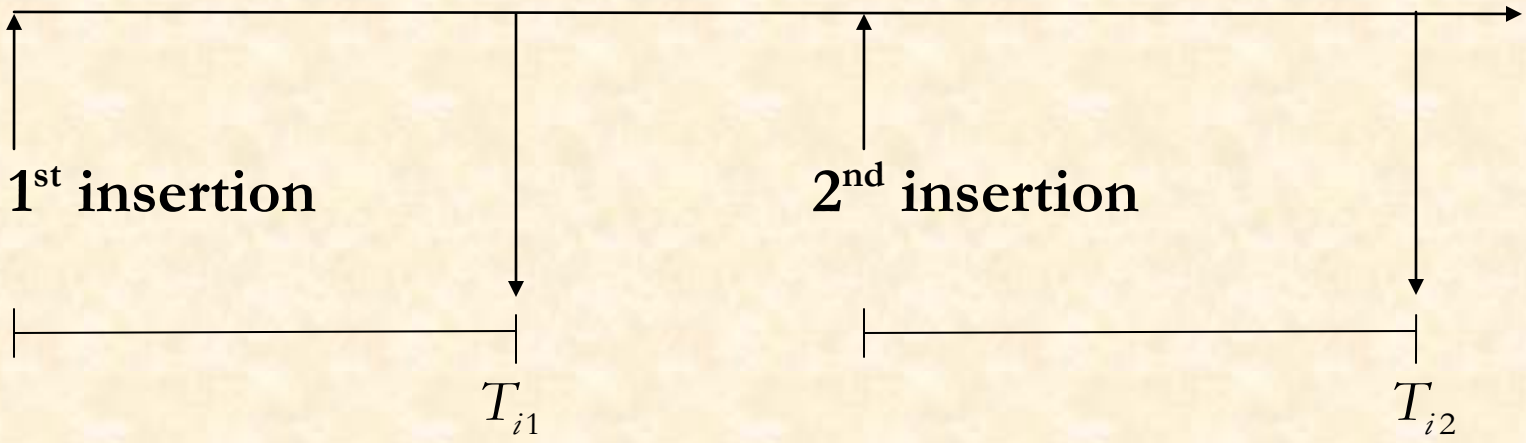
- **ESTIMATION OF SURVIVAL CURVES**
- **COMPARISON OF SURVIVAL CURVES**

Proc LIFETEST is the main engine for nonparametric analysis

ILLUSTRATIVE EXAMPLE 1

A study by McGilchrist & Aisbett (1991) of 38 kidney dialysis patients:

- **Time in days to infection at the catheter insertion point**
- **Two times in each patient** (T_{i1}, T_{i2}) **with possible right censoring at** (U_{i1}, U_{i2})



Data—3 patients

Obs	patient	insert	time	fail	gender	age
1	1	1	8	1	0	28.0
2	1	2	16	1	0	28.0
3	2	1	23	1	1	48.0
4	2	2	13	0	1	48.0
5	3	1	22	1	0	32.0
6	3	2	28	1	0	32.0

age=average age at catheter insertions

```
proc format;
value gender 0='male'1='female';
value insert 1='first' 2='second';
value fail 0='censored' 1='infected';
run;
```

Analysis Variable : time							
insert	gender	fail	N	Mean	Median	Min	Max
first	female	censored	6	58.2	38.0	5.0	149.0
		infected	22	162.2	124.5	7.0	536.0
	male	infected	10	32.8	16.0	2.0	152.0
second	female	censored	10	47.4	24.5	5.0	159.0
		infected	18	119.3	72.0	8.0	333.0
	male	censored	2	6.0	6.0	4.0	8.0
		infected	8	105.8	26.5	9.0	562.0

ESTIMATION OF SURVIVAL CURVES

- **Distribution of infection-free time by insertion and gender**

$S_j(t) = P[T_j > t]$, T_j **- time in days to infection**

- **Nonparametric methods use the data:**

$X_{ij} = \min(T_{ij}, U_{ij})$, $\delta_{ij} = [T_{ij} \le U_{ij}]$ ***j*=1, 2 (stratum),**

$\delta_{ij} = 1$ **if infection time is observed,** $\delta_{ij} = 0$, **otherwise,**

U_{ij} **- censoring time**

Inputs:

- **accumulating count of events up to time *t*,**
  $N_j(t) = \sum_{i=1}^{n} [X_{ij} \le t, \delta_{ij} = 1]$
- **number at risk at time *t*,** $Y_j(t) = \sum_{i=1}^{n} [X_{ij} \ge t]$**.**

Estimators:

$S_j(t)$ **and cumulative hazard** $H_j(t)$ **and are estimated by**

$$\hat{H}_j(t) = \int_0^t \{Y_j(u)\}^{-1} dN_j(u), \; \hat{S}_j(t) = \exp(-\hat{H}_j(t))$$

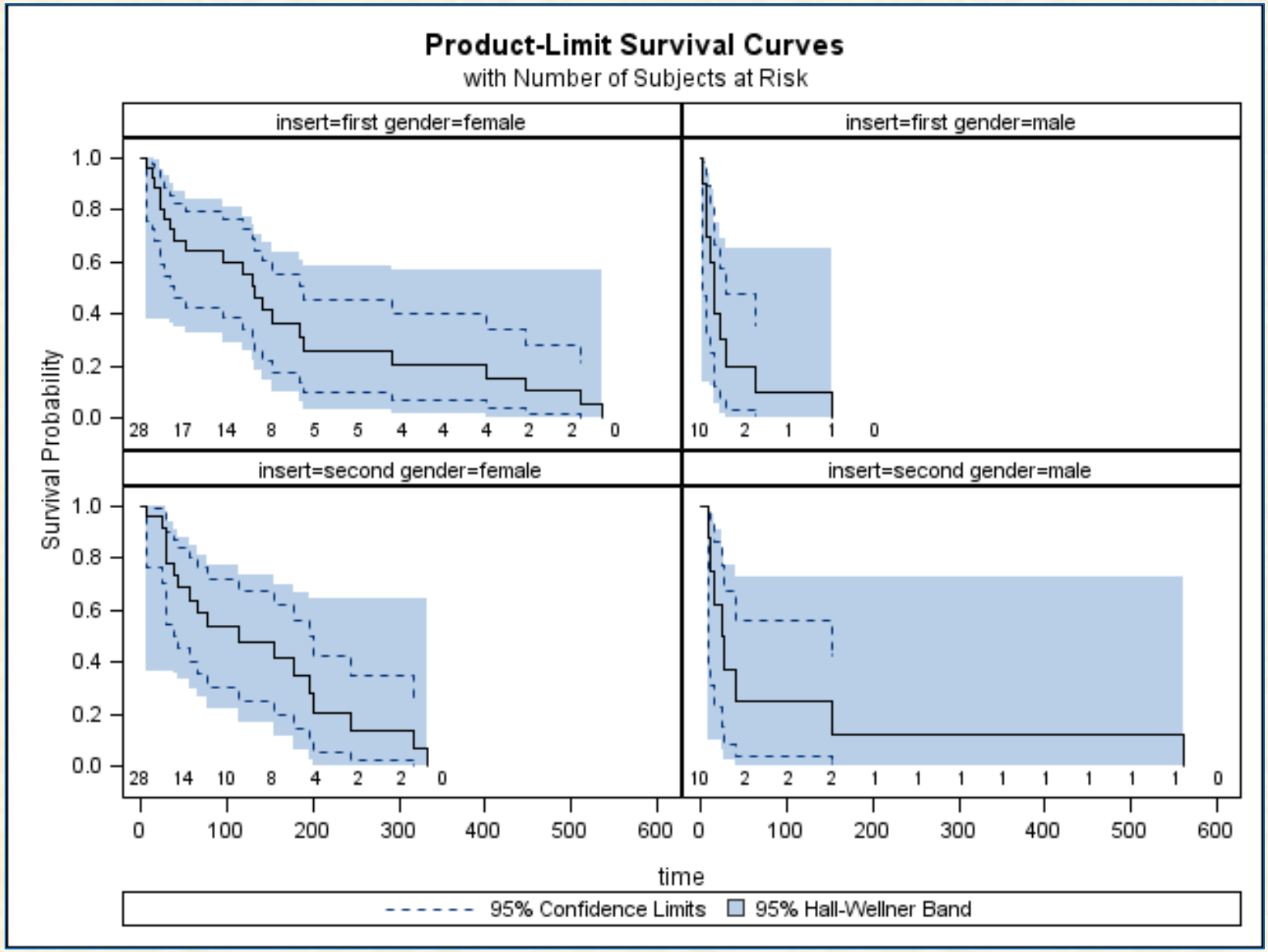
Kaplan-Meier (Product limit) estimator:

$$\tilde{S}_j(t) = \prod_{u \le t} \left(1 - \frac{\Delta N_j(u)}{Y_j(u)} \right)$$

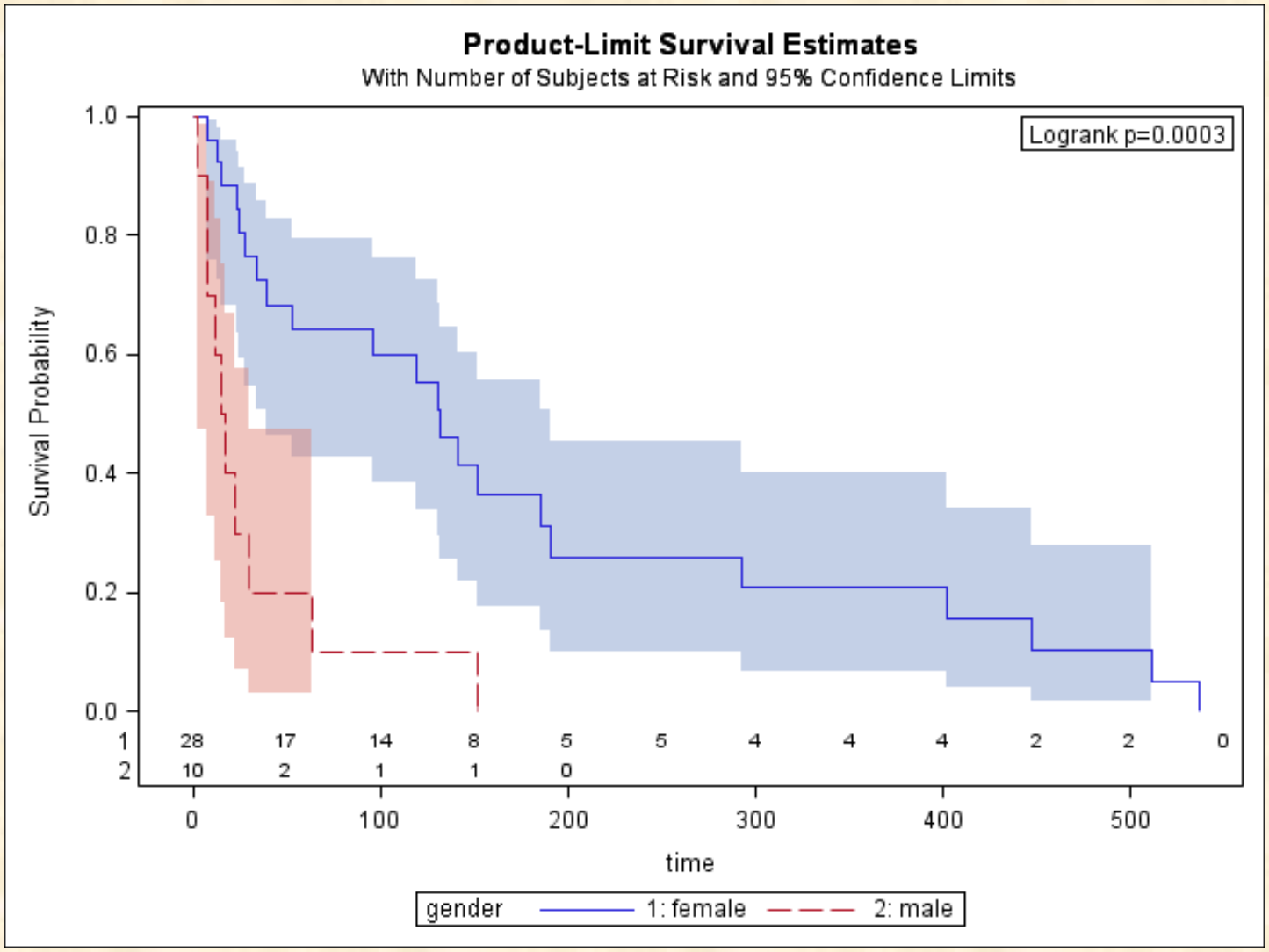
SAS SYNTAX

```
ods graphics on;
proc lifetest data=kidney Method=km
plots=survival(nocensor cb=hw cl strata=panel
                atrisk=0 to 600 by 50);
strata insert gender;
time time*fail(0);
format gender gender. insert insert.;
run;
ods graphics off;
```

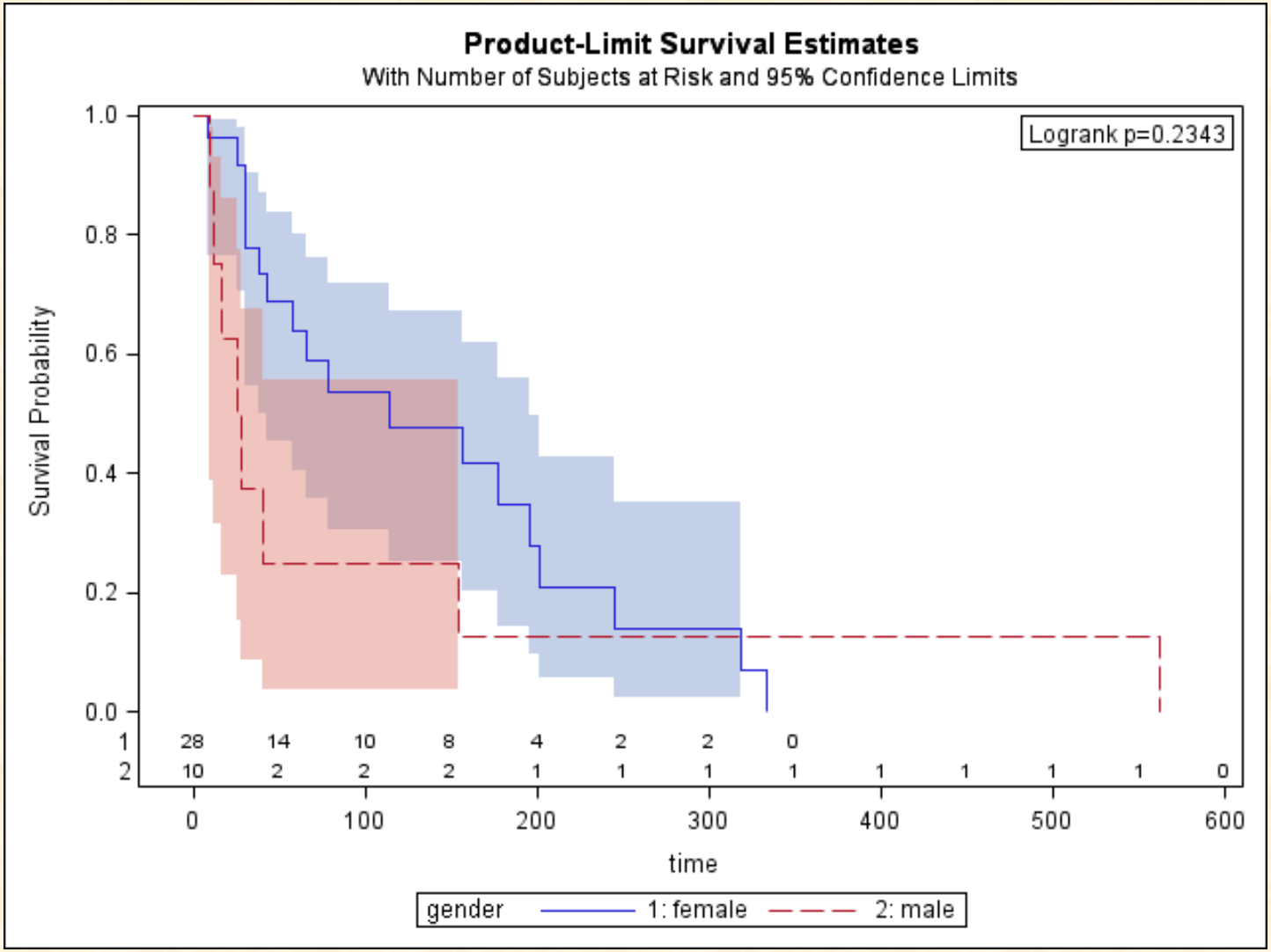
```
nocensor to suppress plotting of censored times
cl for confidence (pointwise) limits
cb=  for confidence bands
atrisk=  to display patients at risk at specified times
strata=panel to display of plots in a panel
```


Product-Limit Survival Curves
with Number of Subjects at Risk
insert=first gender=female
insert=first gender=male
insert=second gender=female
insert=second gender=male
Survival Probability
1.0
0.8
0.6
0.4
0.2
0.0
28 17 14 8 5 5 4 4 4 2 2 0
10 2 1 1 0
28 14 10 8 4 2 2 0
10 2 2 2 1 1 1 1 1 1 1 1 0
0 100 200 300 400 500 600
time
95% Confidence Limits
95% Hall-Wellner Band

Comparisons: by gender at each insertion–First insertion



Second insertion



Some Conclusions

- Logrank test-comparison by gender- obtained from

```
strata insert/group=gender test=logrank;
```

  This is a stratified test (p=.0009). Separately by first, second insertion,

- First insertion– gender effect is significant. Females have longer infection-free duration

- Second insertion– gender effect is not significant

- 95% pointwise confidence intervals are displayed–little overlap for first insertion

II. PARAMETRIC MODELS-ACCELERATED FAILURE TIME MODEL

- **FITTING PARAMETRIC MODELS**
- **ESTIMATION OF PERCENTILES**
- **JOINT MODELING OF INFECTION TIMES**

Procedures LIFEREG and RELIABILITY can be used. Also a new PROC SEVERITY

ACCELERATED FAILURE TIME MODEL

- $\log T = \mu + \sigma\varepsilon$, **where** μ **=location and** σ **=scale are parameters**
- **Covariate effects are modeled by** $\mu = \mathbf{z}'\beta_1$
- **heteroscedasticity by** $\log\sigma = \mathbf{z}'\beta_2$
- **distribution on** ε **independent of z, induces a distribution on** T

Use MLE to estimate (β_1, β_2)

AFT class includes exponential, Weibull, lognormal and loglogistic.

General form: $S(t) = S_0((t/\alpha)^\gamma)$ **where** $\sigma = \gamma^{-1}, \mu = \log\alpha,\ \alpha > 0, \gamma > 0$**,**
and S_0 **is a known survival distribution**

Generalized Gamma (GG) Distribution

- **Additional shape parameter**
- **AFT form:** $\log T = \mathbf{z}'\beta_1 + \sigma_0 Z$ **where** $k = \delta^{-2}, \sigma_0 = \sigma\delta,$

$$Z = \sqrt{k}(\varepsilon - \log k)$$

- **SAS calls** δ **the *shape* and** σ_0 **the *scale* of the GG**

GG returns three special cases:

(1) with δ=0 the log normal. As $\delta \to 0$, Z converges to the standard normal;

(2) with δ=1 the Weibull;

(3) with δ=1 and σ_0=1 the exponential;

FITTING PARAMETRIC MODELS

Assume the within-patient times (T_{i1}, T_{i2}) are independent, making our sample comprise of 76 individual catheter insertions.

```
proc lifereg data=kidney;
class gender;
model time*fail(0)=age gender/dist=gamma;
format gender gender.;
run;
```

For lognormal: $H_0 : \delta = 0.$ **Use dist=gamma noshape1 shape1=0;**

For Weibull: $H_0 : \delta = 1.$ **Use dist=gamma noshape1 shape1=1;**

For exponential: $H_0 : \delta = 1, \sigma_0 = 1.$

Use dist=gamma noshape1 shape1=1 noscale scale=1;

Table 1: Results of fitting parametric AFT models to infection times					
	Maximum likelihood estimate (standard error)				
Parameter	GG	Lognormal	Weibull	Exponential	Loglogistic
Intercept	3.4188 (0.5322)	3.4490 (0.4939)	4.2916 (0.5505)	4.4025 (0.4971)	3.4052 (0.4636)
AGE	–0.0054 (0.0097)	–0.0054 (0.0097)	–0.0042 (0.0103)	–0.0046 (0.0094)	–0.0073 (0.0093)
GENDER-female	1.3830 (0.3283)	1.3269 (0.3261)	0.9655 (0.3248)	0.8853 (0.2871)	1.5526 (0.3265)
Scale	1.1863 (0.1085)	1.1847 (0.1077)	1.1031 (0.1035)	1(fixed)	0.6793 (0.0720)
Shape	–0.0473 (0.3164)	0(fixed)	1(fixed)	1(fixed)	na
–2 log L	197.032	197.053	206.197	207.348	198.532
BIC	218.686	214.375	223.520	220.340	215.855
LM test p-value	na	.887	<.0001	Shape <.001 Scale .316	na

ESTIMATION OF PERCENTILES

- **Given a covariate profile z: the 100(1–p)-th percentile t_p of the event time T is $t_p = \exp(\mathbf{z}'\beta + \sigma w_p)$ where w_p is the corresponding percentile of ε.**
- **Work with $\log \hat{t}_p = \mathbf{z}'\hat{\beta} + \hat{\sigma} w_p$ to obtain 95% CI for t_p**
- **Use PROC RELIABILITY**

See SGF 2010: Paper 252-2010 for an example and syntax

JOINT MODELING OF INFECTION TIMES

- **Allow correlation between** (T_{i1}, T_{i2}) **via a shared frailty model**

  $\log T_{ij} = \mathbf{z}'_{ij}\beta + \nu_i + \sigma\varepsilon_{ij}$ **where** ν_i **is a random effect**

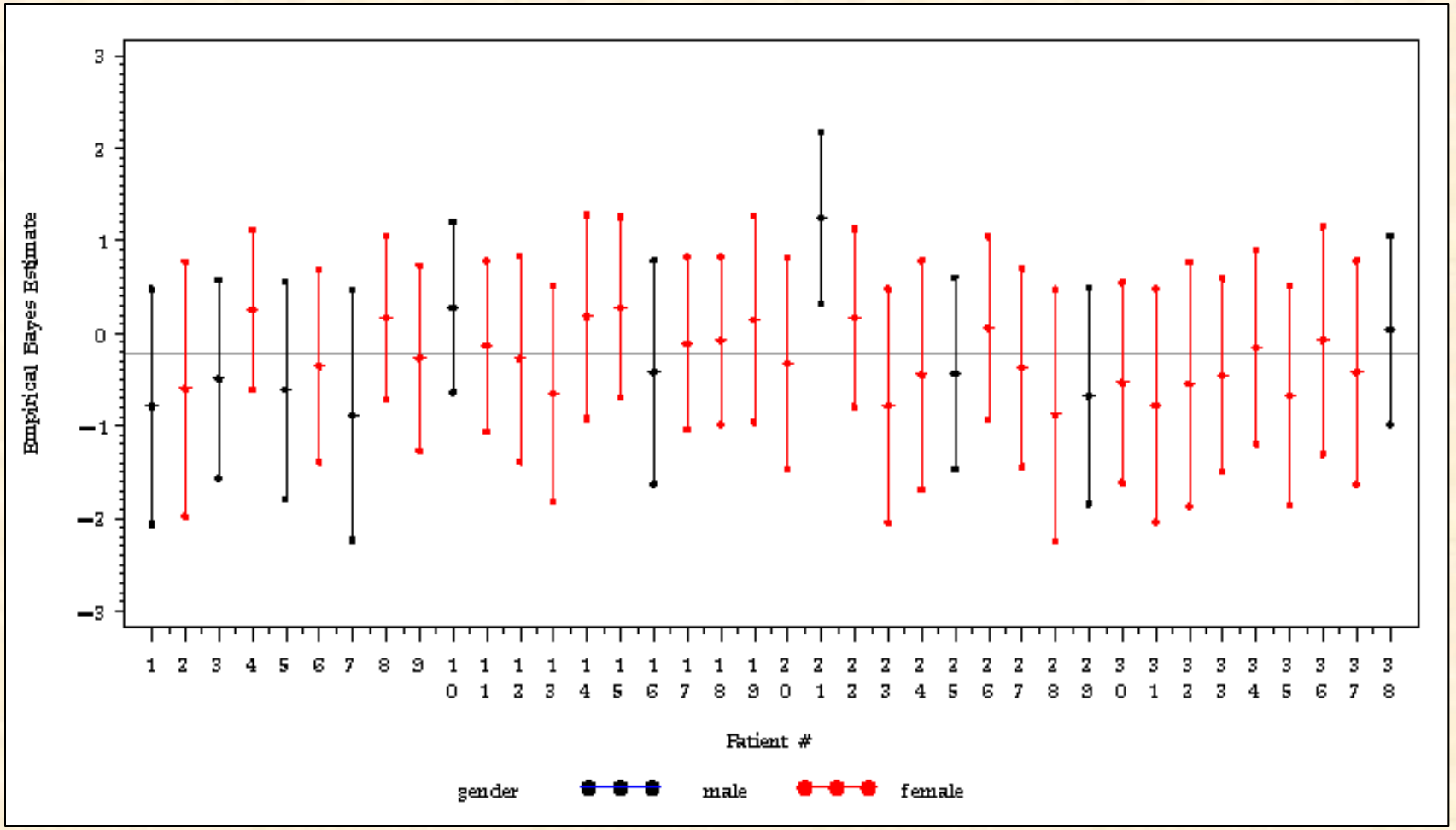
- **Assume** (T_{i1}, T_{i2}) **conditionally independent given** $(\nu_i, \mathbf{z}_{i1}, \mathbf{z}_{i2})$**,**

  parametric distribution for $(\nu_i, \varepsilon_{ij})$ **given** $(\mathbf{z}_{i1}, \mathbf{z}_{i2})$

- **Perform ML estimation of the marginal model using PROC NLMIXED**

Example: Weibull model, ε_{ij} **~extreme-value,** $\nu_i \sim N(-\frac{1}{2}\sigma_\nu^2, \sigma_\nu^2)$.

Get MLE of (β, σ_ν^2)**and empirical Bayes estimates of** ν_i**, that is** $E(\nu_i \mid data)$



LRT for frailty effect p<.01; but the effect appears to be from patient #21

Remarks

- **Other frailty models**

  $\varepsilon_{ij} \sim extreme\text{-}value,\ \xi_i = \exp(\nu_i) \sim Gamma$

  $\varepsilon_{ij} \sim normal,\ \nu_i \sim normal$

  can be fitted by NLMIXED

- **Use MLE from (non-frailty model) LIFEREG to suggest starting values**

- **PROC QLIM in SAS/ETS can fit a joint log-normal model**

  $y_{ij}^* = \log T_{ij} = \mathbf{z}_{ij}'\beta + u_{ij}$ **with** (u_{i1}, u_{i2})**~Normal (0,** ρ, σ_1, σ_2**) and observed range of** y_{ij}^* **is restricted**

  See SGF 2010: Paper 252-2010 for details

SUMMARY

- **Procedures LIFETEST, LIFEREG, PHREG provide tools for comprehensive analyses**
- **PHREG for analyses the Cox model and its extensions**
- **ODS GRAPHICS and PLOT options in above procedures provide exquisite enhancements**
- **BAYES option in LIFEREG, PHREG give additional capabilities**
- **PROC MCMC can be used for more complex survival analyses**
- **Future enhancements could include frailty models, finite mixture models, hyper-prior specification for Bayes models**

ACKNOWLEDGEMENTS

- This talk is based on a presentation at SAS Global Forum 2010, April 11-14, Seattle, WA.

Paper 252-2010: Gardiner JC. "Survival Analysis: Overview of Parametric, Nonparametric and Semiparametric approaches and New Developments"

- Agency for Healthcare Research & Quality for sponsorship of my research under Grant 1R01 HS14206

Thank you for your attendance