Mathematical and Computational Base for Valid Inference with The Two-Stage Clonal Expansion Model -- Conditional Likelihood Method –

T. Nakamura (Chuo Univ.)
D. G. Hoel (Medical Univ. of South Carolina)

Colleagues

S. Kinoshita, Y. Yoshinaga, Y. Yoshida, S. Sato,

S. Nogami, Y. Inoue (Nagasaki Univ.)

ISI Invited Session, IBC 2012, Kobe

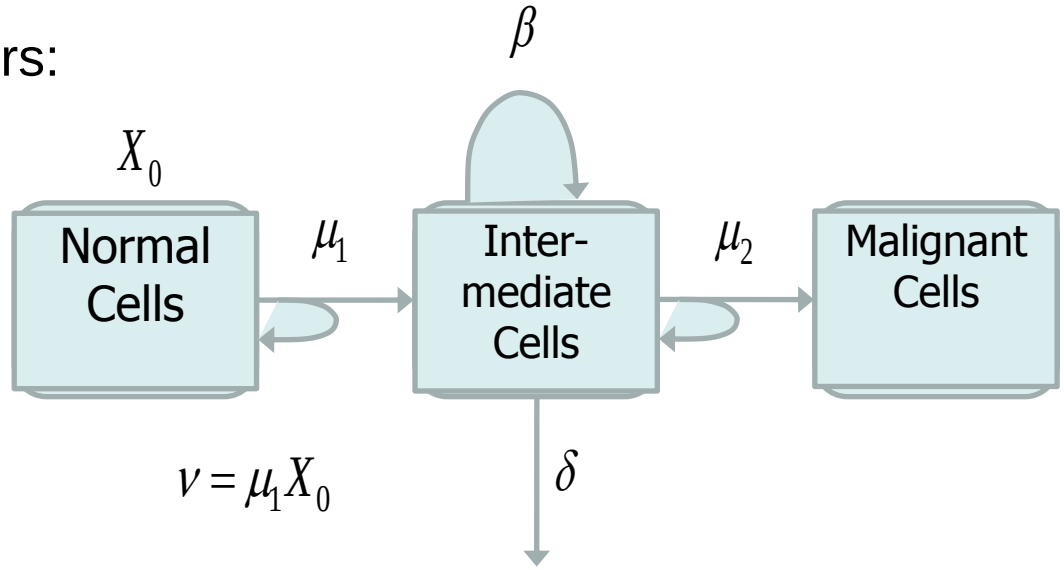
Two-Stage Model was proposed by Moolgavkar-Venzon-Knudson (1979).

The model involves four parameters:
μ_1, β, δ, and μ_2.

A normal cell is transformed into an intermediate cell at a constant rate μ_1 per unit time per cell.

Each intermediate cell is subject to three events:

birth, death or mutation.



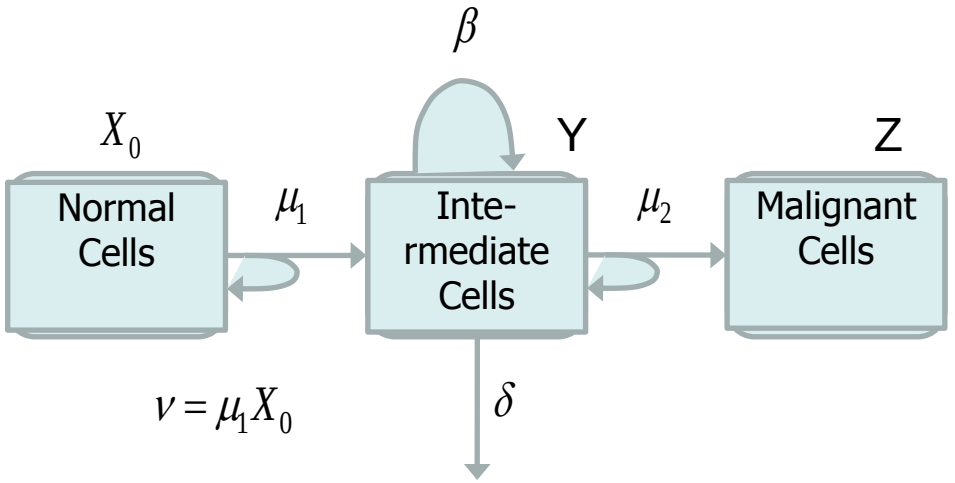
Specifically, each intermediate cell undergoes clonal expansion by dividing into two intermediate cells at a rate of β per unit time, death at a rate of δ per unit time, or division into one intermediate cell and one malignant cell at the rate of μ_2 per unit time.

Once a malignant cell has developed, it will replicate into a detectable tumor in a relatively short period of time; therefore, the formation of one malignant cell is usually associated with the development of detectable cancer (Moolgavkar et al.)

Hornsby (Lancet 2007):

Its biggest triumph lies in its stochastic representation of clonal growth.

Let X_0, Y(t), Z(t) denote the number of normal, intermediate and malignant cells, at t respectively. X_0 is a constant.



Endpoint:
Appearance of first malignant cell.

$$S(t) = \Pr\{Z(t) = 0 \mid Y(0) = 0, Z(0) = 0\}$$

: Probability of no malignant cell at t

2. Moolgavkar etal.(1988) obtained differential eq. for *S(t)* using Prob. Generating function of Y(t), Z(t)

3. **Portier-Kopp-Sherman** (PKS 1994) found an exact solution to the differential eq. and a closed form of *S(t)*. Annette will describe this.

4. Problem is that $\nu,\ \beta,\ \delta,\ and\ \mu_2$ are jointly **un**identifiable with survival data.

SOME IDENTIFIABLE SET OF PARAMETERS ARE PROPOSED:

(A) Hanin et al. and PKS (1996-7)

$$R^2 = (\beta + \delta + \mu_2)^2 - 4\beta\delta$$

$$\Psi = \beta - \delta - \mu_2 \quad (\textit{net-proliferation rate})$$

$$\eta = \nu / \beta$$

$$\lambda(t) = \eta(R + \Psi)\left[\frac{1}{2} + \frac{-R\exp(-Rt)}{R - \Psi + (R + \Psi)\exp(-Rt)}\right]$$

(A*) Heidenreich et al (Risk Anal. 1997)

$$\rho = \nu\mu_2 \quad (\textit{overall mutation rate})$$

$$\Psi = \beta - \delta - \mu_2$$

$$q = 2\beta\mu_2 / (\Psi + R)$$

$$\lambda(t) = \frac{\rho[\exp\{(\Psi + 2q)t\} - 1]}{q[\exp\{(\Psi + 2q)t\} + 1] + \Psi}$$

Likelihood based on these hazards will be termed "**Original likelihood**"

5. Problems with the Original likelihood

a) **Sherman and Portier** (Risk Analysis, 1997.):

Restrictions on Parameters

Estimated parameters are quite different with the different restrictions.

However, predicted incidence curves are nearly identical to each other.

Estimability:

The original likelihood function was extremely flat in the region of the optimum and considerable flexibility in determining MLE.

Interpretation:

Caution should be used in interpreting the findings from fitting two-stage models to survival data alone.

b) **Kopp and Portier** (Fund. Appl. Toxic. , 1991) :

Model fitness and model selections

Radically different models could fit survival data equally well.

Goodness-of-fit testing or likelihood comparisons are useless in model selections

~~~~~~

c) **Kopp** (Stat. Meth. Med. Res. 6, 1997)

I would like to leave this for Annette , the next speaker.
~~~~~~

5. Problems with the Original likelihood continued

d) Portier and Marsi (*Statistical research needs in mechanistic modeling for carcinogenic risk assessment*, Stat. Meth. Med. Res. 6, 1997)

- **Caution**: Flat likelihood surface resulted in large variance estimates and a lack of confidence in the biological meaning of the resulting values.

  Additional assumptions on parameters or additional data such as $X=10^8$ or *labeling index* obtained from separate animals biased estimates.

- **Require**: Statistical methods for the estimation of parameters (e.g. new likelihoods)

e) Hornsby (Lancet 8, 1030~38, 2007)

Caution

- Different studies that have used slightly different methods or datasets have yielded different conclusions.
- The carcinogenesis models need to be carefully applied, and the conclusions they yield should be treated with caution.

f) Heidenreich et al. (Rad Env Biophys,253–264, 2000)

Subjects and Methods:

Three carcinogenesis modeling groups (Heidenreich & Paretzke, Germany; Brugmans & Leenhouts, The Netherlands; Little, UK) separately applied TSCE model to one data set of lung tumors in rats exposed to radon (n=6574, d=406).

Results:

Although the three groups used one carcinogenesis model, different model assumptions and/or different methods of finding the "best fit" could result in different descriptions of experimental data.

Conclusion:

Even for the same data and comparable mathematics, different solutions can be found. Additional data will be needed to understand the nature of the radiation action within the framework of the TSCE model.

6. Nakamura and Hoel: Conditional Likelihood Method, 2001

Put δ=0, then $$\lambda(t \mid \delta = 0) = \nu^* \mu_2^* \frac{1 - \exp[-(\beta^* + \mu_2^*)t]}{\mu_2^* + \beta^* \exp[-(\beta^* + \mu_2^*)t]}$$

* denotes these are conditional.

Question: Much simpler! But, *what are the biological meaning of them?*

Answer: *It holds exactly that* $\psi = \beta^* - \mu_2^*$, $\rho = \nu^* \mu_2^*$, $\eta = \nu^* / \beta^*$

and approximately that $\beta^* \cong \psi \cong \beta - \delta$

Dose-rate estimation: Log-linear model for all parameters

Original parameters

$$\log \rho = \rho_0 + \rho_D D$$
$$\log \Psi = \Psi_0 + \Psi_D D$$
$$\log \eta = \eta_0 + \eta_D D$$

Conditional parameters

$$\log \nu^* = \nu_0^* + \nu_D^* D$$
$$\log \beta^* = \beta_0^* + \beta_D^* D$$
$$\log \mu_2^* = \mu_{20}^* + \mu_{2D}^* D$$

Transformations

Since $\rho = \nu^* \mu_2^*$ and $\Psi \cong \beta^*$,

$$\rho_0 = \nu_0^* + \mu_{20}^*, \rho_D = \nu_D^* + \mu_{2D}^*,$$

$$\Psi_0 \cong \beta_0^* \text{ and } \Psi_D \cong \beta_D^*$$

Thus, ρ_0, ρ_D , Ψ_0 and Ψ_D are obtained from the conditional likelihood without any additional assumption. Three models are biologically interesting:

Initiation-model: Ψ_0, ρ_0, η_0 plus ρ_D / $\nu_0^*, \beta_0^*, \mu_{20}^*$ plus ν_D^*

Promotion-model: Ψ_0, ρ_0, η_0 plus Ψ_D / $\nu_0^*, \beta_0^*, \mu_{20}^*$ plus β_D^*

Full-model: $\Psi_0, \rho_0, \eta_0, \Psi_D, \rho_D, \eta_D$ / $\nu_0^*, \beta_0^*, \mu_{20}^*, \nu_D^*, \beta_D^*, \mu_{2D}^*$

7. Data to analyze

Kociba et al. (1978, 1990) dealt with rats maintained for two years on diets supplying 0, 1, 10 and 100pg TCDD/ kg. The data has been used by US-EPA for risk assessment The endpoint is the presence of adenoma or hepatocellular carcinoma on the day of death, the other deaths being treated as censored.

JANUS data is obtained from mice irradiated to γ or *neutron* in Argonne National Laboratories between 1970 and 1992. The experiment was conducted to study the long-term effects of an external wholebody irradiation. The mice were examined histologically at time of death.

Two Convergence criteria for the Newton-Raphson iteration for MLE

A-mode, or **true, convergence**
- The observed information I >0
  Operationally,
  The diagonal elements of I^{-1} >0
- Norm $||U||$of Score vector <0.0001=1E-4
- Norm $||\Delta\beta_n||$of increment <0.00001=1E-5

B-mode, or **false, convergence**
- A-mode convergence fails to hold.
- The diagonal elements of I^{-1} >0
- $||U||$ is minimum among the first 100 iterations.

When a convergence is of A-mode, the likelihood surface has a clear peak at MLE that is attained usually in a few iterations.

As for B-convergence, the likelihood surface is flat and the iteration with different initial values tend falsely to converge to different MLE.

Example 1: **Kociba** data: n=205, d=31.

Conditional likelihood for P-model: A-convergence

$\|U\| = 6.6\text{E-}5, \|\Delta\beta_n\| = 7\text{E-}5, \ell = -202.1307$

	$[1]\nu_0^*$	$[2]\mu_{20}^*$	$[3]\beta_0^*$	$[4]\beta_D^*$
Est.	-5.006	-29.82	-3.330	0.0036
SE	0.3346	7.373	0.308	0.0006

I^{-1}

[1]	0.112	1.081	-0.048	-1.3E-4
[2]	1.081	54.36	-2.27	-2.0E-3
[3]	-0.048	-2.27	0.095	8.4E-5
[4]	-1.3E-4	-2E-3	8.4E-5	3.6E-7

Conditional likelihood for I-model: B-convergence

$\|U\| = 2.5\text{E-}3, \|\Delta\beta_n\| = 5\text{E-}4, \ell = -202.3067$

	$[1]\nu_0^*$	$[2]\mu_{20}^*$	$[3]\beta_0^*$	$[4]\nu_D^*$
	1.588	-25.72	-3.933	0.0265
	1.744	0.45	0.139	0.004

	3.042	−0.1800	−0.2310	−1.7E−3
	−0.18	0.1990	−1.5E−3	−5.0E−6
	−0.231	−1.6E−3	1.9E−2	6.2E−5
	−1.7E−3	−5.0E−6	6.0E−5	1.5E−5

Original likelihood for P-model: B-convergence

$\|U\| = 44.66, \|\Delta\beta_n\| = 8\text{E-}6, \ell = -202.1307$

Therefore, MLE were obtained from the conditional MLE by transformation

	Ψ_0	ρ_0	η_0	Ψ_D	η_D
Est.	-3.330	-34.83	-1.676	0.0036	-0.0036
SE	0.308	7.526	0.5511	0.0006	0.0006

Conditional likelihood converges in A-mode for P-model but in B-mode for I-model, indicating Initiation effect is unlikely.

TCDD increases the promotion rate by exp(0.0036) per pg.

Original likelihood fails to converge.

Kociba on Promotion model: Conditional vs Original

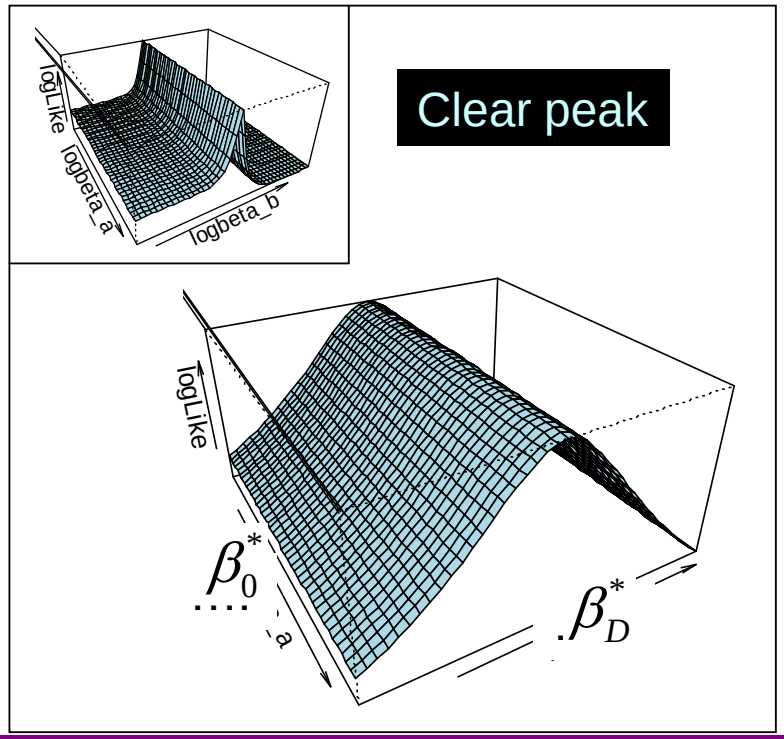
Conditional Likelihood :A-convergence

Fix $\nu_0^* = 5.006,\ \mu_{20}^* = -29.82$

x-axis : $\beta_0^* = -3.33\ (-4.33 \sim -2.33)$

y-axis: $\beta_D^* = 0.00355\ (0.00255 \sim 0.00455)$

z-axis: logL = -696.5 ~ -202.13



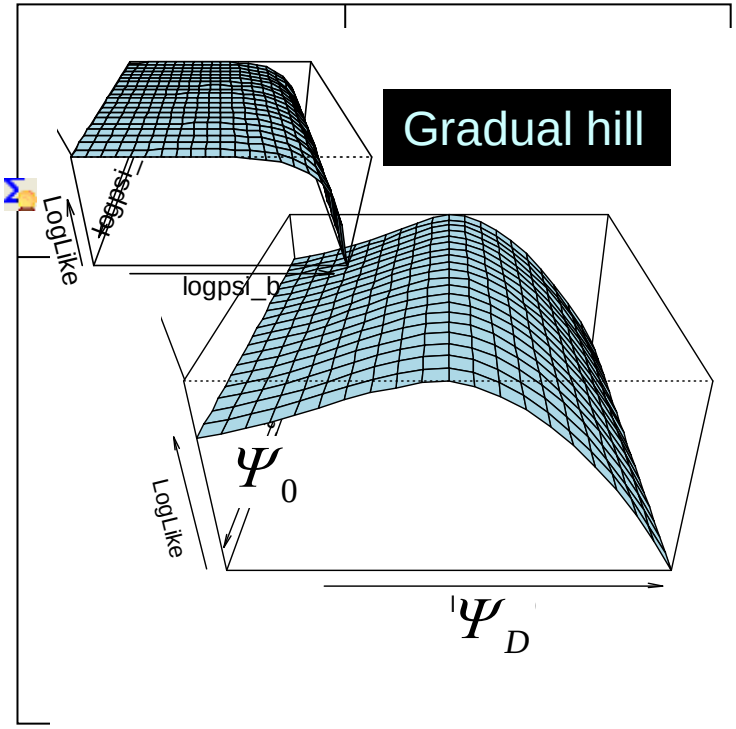
Original Likelihood : B convergence

Fix $\rho_0 = -34.83,\ \eta_0 = -1.676,\ \eta_D = -0.00355$

x-axis : $\Psi_0 = -3.33\ (-4.33 \sim -2.33)$

y-axis: $\Psi_D = 0.00355(0.00255 \sim 0.00455)$

z-axis: logL = -1599.3 ~ -202.13



The left graph shows the likelihood surface for which MLE was clearly attained by N-R algorithm. The 3 base parameters and 1 dose-rate parameter, β_D, are used. MLE and the range of each axis are shown. There is a clear peak for β_D. The right graph shows the likelihood surface for which N-R algorithm fails to converge in 100 iterations.

Conditional Likelihood for Kociba: Promotion vs Full

Promotion : A-convergence

Fix $\nu_0^* = -5.006,\ \beta_0^* = -3.330$

x-axis : $\mu_{20}^* = -29.82\ (-39.82 \sim -19.82)$

y-axis: $\beta_D = 0.0036(0.0026 \sim 0.0046)$

z-axis: $log\mathrm{L} =$ -447.5 ~ -202.1

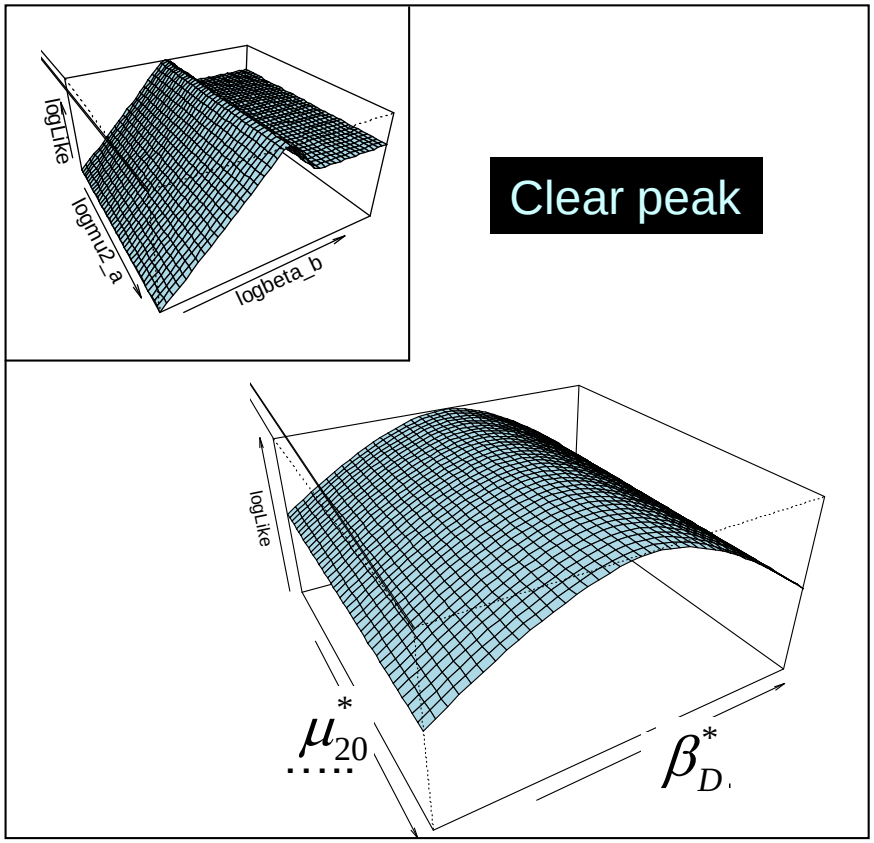
Full : B-convergence

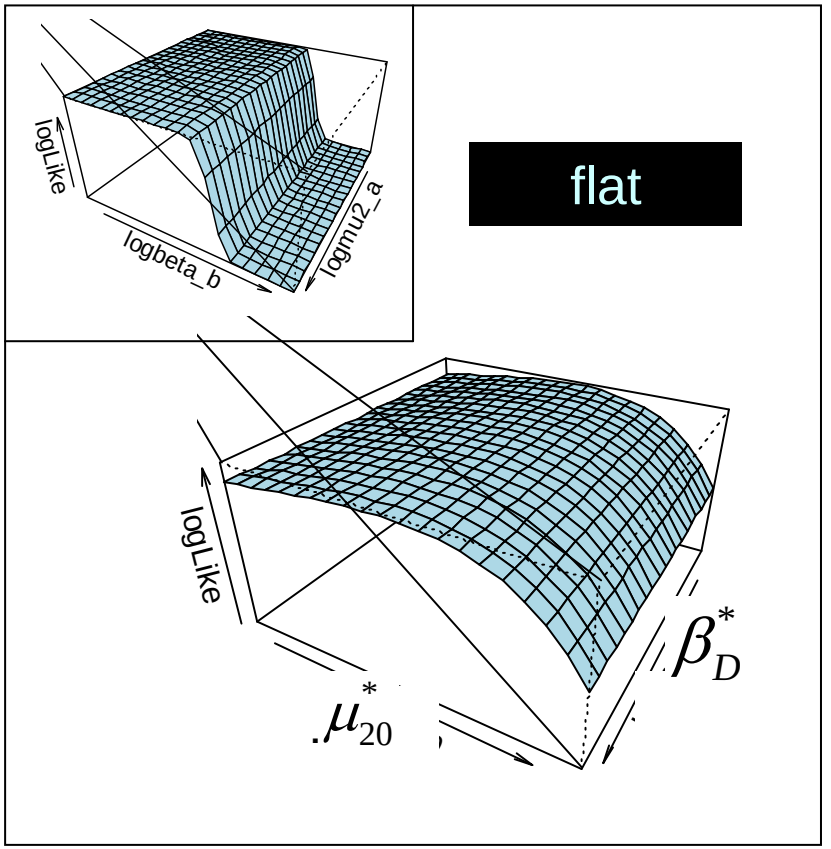
Fix $\nu_0^* = -4.037\ ,\ \beta_0 = -3.521$

x-axis : $\mu_{20}^* = -26.24\ (-28.74 \sim -23.74\ )$

y-axis: $\beta_D^* = -0.0063(-0.0065 \sim -0.0060)$

z-axis: $log\mathrm{L} =$ -461.8 ~ -202.1





8. Simulation Studies for Conditional Likelihood

The asymptotic unbiasedness of the conditional parameters in the loglinear model have been confirmed by a series of simulation studies with n=500.

Yoshida, Sato, Nakamura. *Small sample performance of the two-stage carcinogenesis model*, 56th ISI, Lisboa, 2007.

Nogami, Yoshida, Nakamura: *On the reliability of MLE in the TSCE carcinogenesis model.* Int. Nat'l Conf. on Stat. Meth. for Risk Analysis, ISI Satellite Conference, Lisboa, 2007.

Yoshida, Sato, Takatsuji, Nakamura: *Statistical Inference on Two-Stage Clonal Expansion Model*, ISI-CRA3, Greece, 2009.

Yoshida, Sato, Nakamura, Hoel: Conditional likelihood approach for a two-stage clonal expansion model, *Risk Management* (ed. Jordano and Sousa), 203-224, 2010.

Nakamura and Hoel: Conditional Likelihood for thee Two-stage Clonal Expansion Model, unpublished paper, 2003~

To the best of our knowledge, No simulation studies have been reported that confirmed the validity of the Heidenreich's method for estimation based on the original likelihood.

Example 2 **Nakamura and Hoel:** Environmetrics, 14, 2003

Conditional likelihood for **JANUS**: Endpoint is death due to cancer n=7402 d= 4133

γ–rays **Full** model **A**-Convergence $\|U\|=$ 1E-5, $\|\Delta\beta_n\|=$ 2E-8, ℓ=-29446.81

	$[1]\nu_0^*$	$[2]100\nu_D^*$	$[3]\beta_0^*$	$[4]\,100\,\beta_D^*$	$[5]\mu_{20}^*$	$[6]\,100\,\mu_{2D}^*$	ρ_0	$100\rho_D$
Est.	–4.952	0.741	–4.845	–0.0384	–12.44	–0.306	–17.39	0.435
SE	0.0802	0.101	0.028	0.0068	0.101	0.112	0.118	0.023

I^{-1}

[1]	6.43E-3	-1.32E-1	-1.49E-3	-2.18E-3	1.37E-5	2.58E-4
[2]	-1.32E-1	1.03E+0	-1.67E-1	4.34E-3	-5.33E-3	8.83E-5
[3]	-1.49E-3	-1.67E-1	7.67E-4	-4.67E-4	8.80E-6	-9.24E-5
[4]	-2.18E-3	4.34E-3	-4.67E-4	1.29E-4	-1.25E-4	-2.97E-6
[5]	1.37E-5	-5.33E-3	8.80E-6	-1.25E-4	1.25E-6	1.58E-6
[6]	2.58E-4	8.83E-5	-9.24E-5	-2.97E-6	1.58E-6	4.57E-7

Original Likelihood: Initial values were obtained from the conditional MLE

γ–rays **Full** model **B-convergence**: $\|U\|=0.002$, $\|\Delta\beta_n\|<0.00001$, $\ell=-29446.88$

	η_0	$100\eta_D$	Ψ_0	$100\Psi_D$	ρ_0	$100\rho_D$
Est.	–0.108	0.779	–4.845	–0.0384	–17.39	0.435
S.E	0.101	0.117	0.0277	0.0068	0.118	0.023

I^{-1} *is not obtained in B-convergence because even MLE's are unreliable*

JANUS on Initiation Model: Conditional vs Original

Conditional Likelihood : A convergence

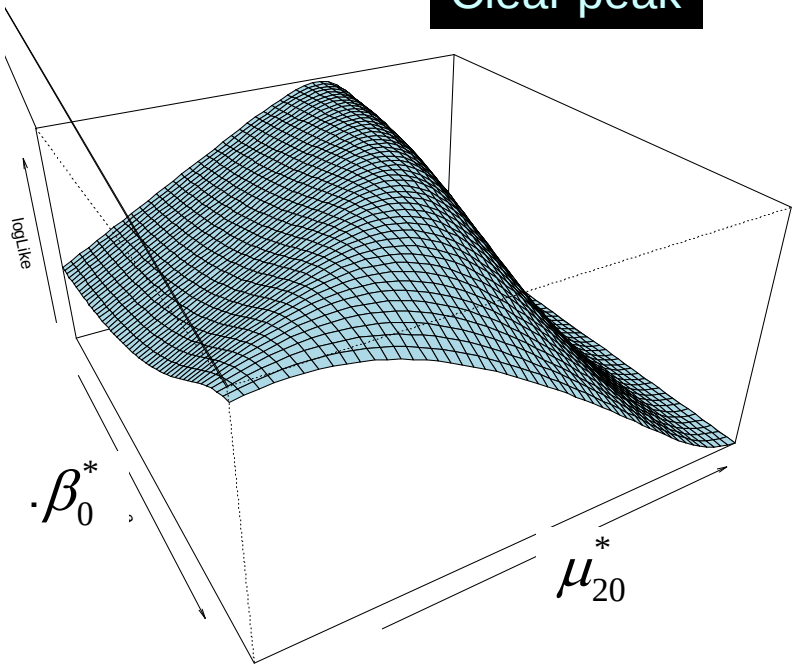
Fix $\nu_0^* = -4.712,\ \nu_D^* = 0.00328$

x-axis : $\mu_{20}^* = -12.49$ (-22.49 ~ -2.49)

y-axis: $\beta_0^* = -4.9$ ($-5.9 \sim -3.9$)

z-axis: $log\mathrm{L}$ = - 29469.16 ~ - 105231.0

Clear peak



Original Likelihood : B convergence

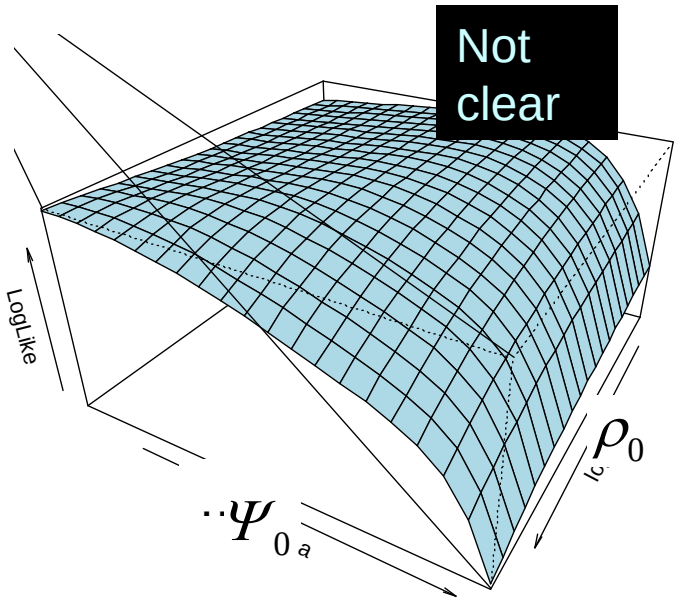
Fix $\rho_D = -0.00328,\ \eta_0 = 0.1869,\ \eta_D = 0.00328$

x-axis : $\rho_0 = -17.2$ ($-7.2 \sim -27.2$)

y-axis: $\Psi_0 = -4.9$ (-5.9 ~ -3.9)

z-axis: $log\mathrm{L}$ = - 431730.5 ~ - 29457.0

Not clear



Example 3 **Heidenreich et al**.: Rad. Res. 166, 2006.

Original likelihood for JANUS: Endpoint is death due to lung cancer.

γ rays and Neutron , n=15975, d=3918 Probably B-convergence

Parameters: $\nu(=X\mu_1)$, $\Psi=\beta-\delta-\mu_2$, $q=2\beta\mu_2/(\Psi+R)$, $Y_0=\nu(0)\mu_2(0)$ and t_{lag}

Dose Effect Model: $\nu(d)=\nu(0)\{1+\nu_{lin}De^{-\nu_{\exp}D}\}$, $\Psi(D)=\Psi_0+\Psi_{level}\{1-e^{-(\Psi_{lin}/\Psi_{level})D}\}$

Final Model

Table 3: MLE for the parameters

γ rays	Male	Female
ν_{lin}	0.076(0.05, 0.01)	0.14 (0.07, 0.24)
Ψ_{lin}	0.75 (0.49, 1.01)	1.2 (0.37, 2.0)
Ψ_{level}	22 $(3,\infty)$	60 $(1,\infty)$
Neutron	Male	Female
ν_{lin}	0.72 (0.29, 1.13)	1.90 (1.20, 2.82)
$\nu_{\exp}$	0.14 (0.05, 0.21)	0.11 (0.07, 0.17)
Ψ_{lin}	3.7 (0, 12)	1.2 (5.0, 11)
Ψ_{level}	1.6 $(0,\infty)$	8.2 (1.9, 15)

Uncertainties

1. The q is not defined.
2. $e^{-(\Psi_{lin}/\Psi_{level})}$ misses D.
3. No MLE of ν_0, Ψ_0, q, $\rho(0)$
4. No Convergence status
5. No I^{-1}
6. Ψ_{level} is not significant
7. Some more below

Uncertainties continued

"Parameter uncertainties are calculated using the profile likelihood technique and the Fisher information matrix of second derivatives. These estimations were done with the function minimizer MINUIT from CERN (17)."

The statement is statistically unclear. No quantitative results are shown.

"As a means for judging the quality of fits, for each group the number of expected cases is calculated by summing over all animals the cumulative hazard of each animal during the follow-up." in Methods.

Portier et al. point out that model fitness based on the expected and observed number of events is useless in model selection. I think this is a common knowledge among applied statisticians too.

"The parameters in these equations are estimated, in addition to those that affect the background only: the product $Y_0 = \nu(0)\mu_2(0)$ of the spontaneous initiation and transformation rates, and the stochasticity parameter *q,* which can give a leveling of hazard to Y_0/q at high age. ……To prevent nonidentifiability, Y_0 is fixed to an arbitrary number."

$Y_0 = \rho_0$ *is estimated with small SE in the Conditional Likelihood method.*

"Including the nonlinearity ν_{exp} for γ rays gives an improvement in the case of male mice, but not a significant one. Contrary to this, for *neutrons* it does provide a significant improvement (not shown). "

Why not shown?

"The estimated lag time is about 250 days in the models for the γ rays and about 170 days in the models for neutrons. It has been verified that the uncertainties of this lag time are large by fixing it to selected values. The standard tools of MINUIT cannot be used for this purpose, because the ages are given discretely, in days. Therefore, the deviances depend on the lag time in a way that is too coarse to draw conclusions from the matrix of second derivatives (Fisher information matrix), and the profile-likelihood calculation also becomes unreliable. The true lag time of the controls cannot depend on the radiation quality. But the lag time can best be estimated from time since exposure, so the statistical power is coming mostly from the groups with high dose."

Too self-involved for me to follow. How was the lag time treated in the results?

For estimation, Heidenreich assumes a dose-rate effect only for $\nu = X\mu_1$ but not for μ_2. This assumption is against basic biology, making the biological meaning of all parameters unclear. Fortunately, the model $\nu(d) = \nu(0)\{1+\nu_{lin}D\}$ is equivalent with our model $\log\nu(D) = \nu_0^* + \nu_D^* D$ when $\nu_{lin}D$ is small. Then, our simulation study have revealed that MLE of $\nu_{lin}\,(\cong \nu_D)$ is actually that of $\rho_D = \mu_{1D} + \mu_{2D}$

9. Dynamic Search Algorithm for valid initial values.

Example: Kociba data: We failed to have Newton-Raphson algorithm converge for $\log\beta^* = \beta_0^* + \beta_D^* D$, but succeeded for $\log\beta^* = \beta_0^* + \beta_{\log D}^* \log(1+D)$ for which

MLE's are $\nu_0^* = -3.94$ $\mu_{20}^* = -20.73$ $\beta_0^* = -3.98$ and $\beta_D^* = 0.075$

With those values as initial values, the Dynamical Searcher searched valid initial values for the model $\log\beta^* = \beta_0^* + \beta_D^* D$

Dynamical Searcher starts

> source("E:¥¥Two Stage Model¥¥TwoStageProgram_New_SKinoRev4.R")

- TSM_Analysis(1)
- P = range <- c(-3.99508, -20.70492, -4.0, 0, 0, 0.075),

Aftermath : reduction x = 5, move m= 6

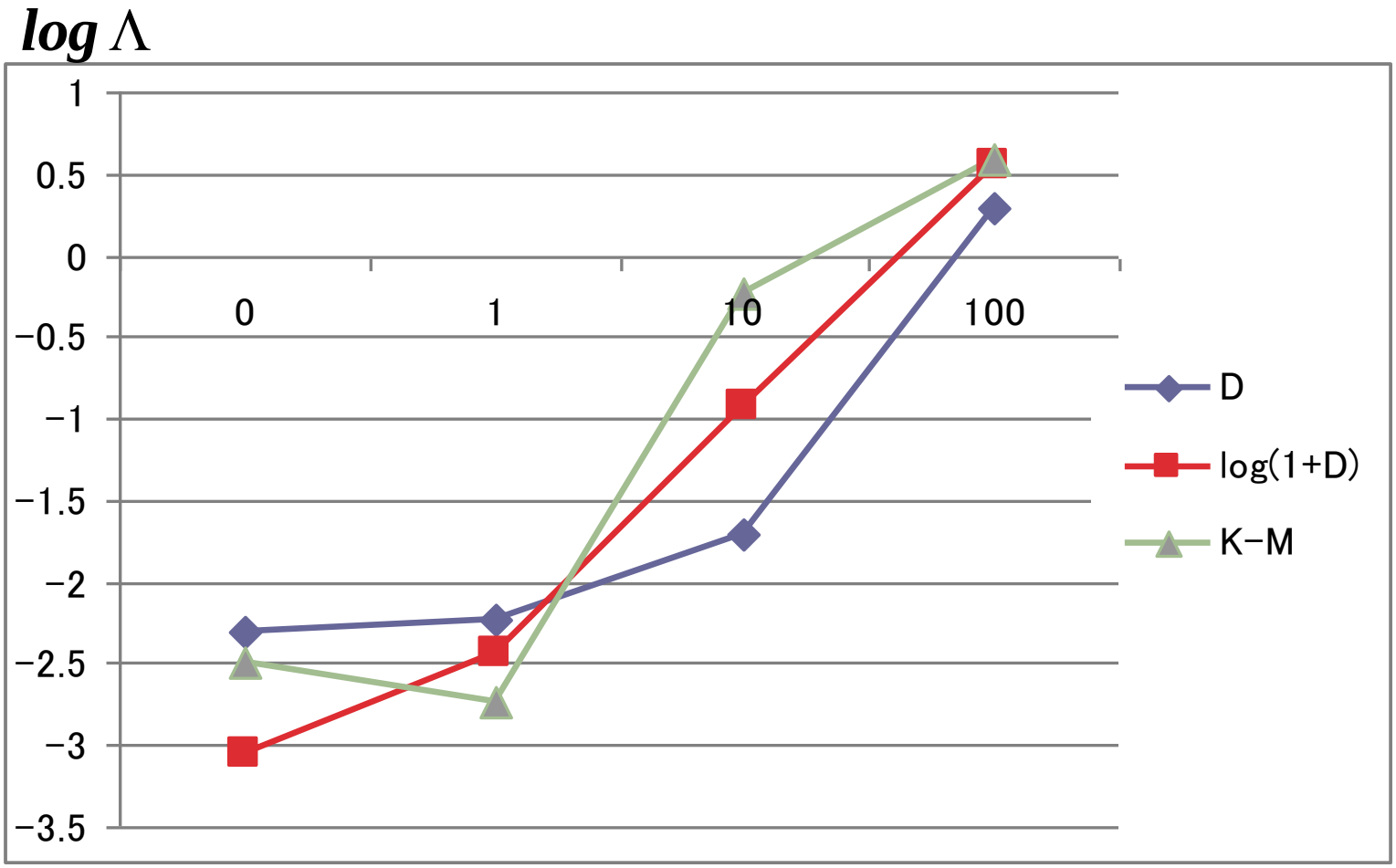
proc = m m x m x m x m x m x

	ℓ	ν_0^*	μ_{20}^*	β_0^*	[4]	[5]	β_D^*
1	-213.9755	-4.994990	-29.98459	-3.324547	0	0	0.003549965
2	-202.3071	-5.003259	-29.96455	-3.324718	0	0	0.003544914
3	-202.2230	-5.012780	-29.95802	-3.324927	0	0	0.003567010
4	-202.4812	-5.010026	-29.95496	-3.324858	0	0	0.003563940
5	-202.1425	-5.011516	-29.95247	-3.324863	0	0	0.003564916

Appropriate Initial Values :	ν_0^*	μ_{20}^*	β_0^*	β_D^*
	-5.01152	-29.95247	-3.32486	0.00356

[1] " Finished "

log Cumulative Hazard estimated by K-M, log (1+D) and D



K-M curve, or an observed result, shows *U*-shaped dose-response.
But both *D* and *log* models are monotone, so fitting is not very well.
D-model fits better in low doses and *log*-model better in higher doses.
We will add a simple nonlinear term to obtain a better fit.

Conclusion

To respond to the caution and requests by the Portier school and Hornsby, we have developed the Conditional Likelihood Method, the Dynamical Searcher for MLE, performed extensive simulation studies to confirm that the method is valid, reliable and flexible with small sample sizes.

Heidenreich et al. (Rad. Env. Bioph. 2012) also shows little critical information. We are concerned about their interpretation of data assuming convenient conditions, too brief account of inference methods, not presenting the convergence status and the information matrix, because the interpretation of the results from the TSCE model heavily depends on those information as well as MLE itself.

The Conditional Likelihood Method is indeed what the Portier school and others requested and merits further study, with the goal of establishing a standard inference method for more reliable and broader applications of this useful model.