Estimating parameters from CFSE data

Rob J. de Boer
Theoretical Biology, UU

Ongoing work with:

Rustom Antia, Vitaly Ganusov, Sergei Pilyugin
&
Alan Perelson

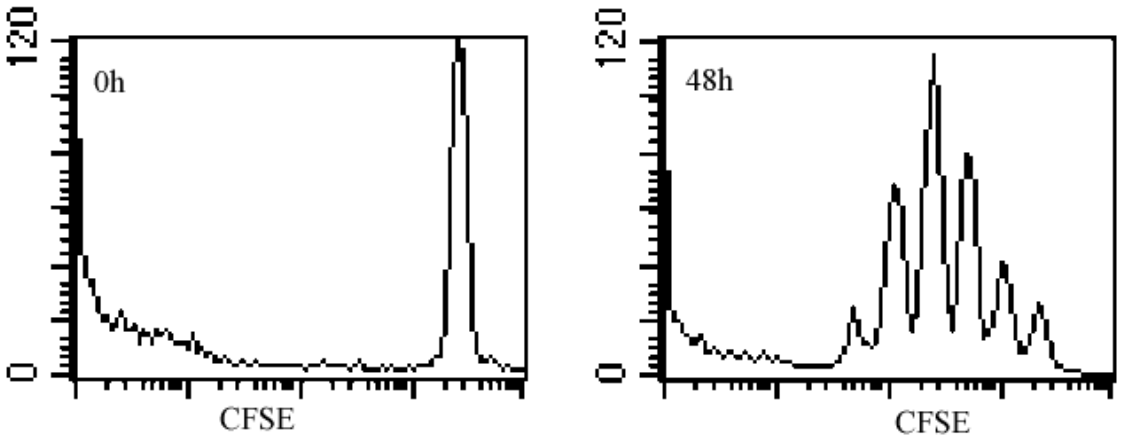
CFSE

- Carboxy Fluorescent Succinimidyl Ester

$\rightarrow$ Lyons & Parish, JIM, 1994

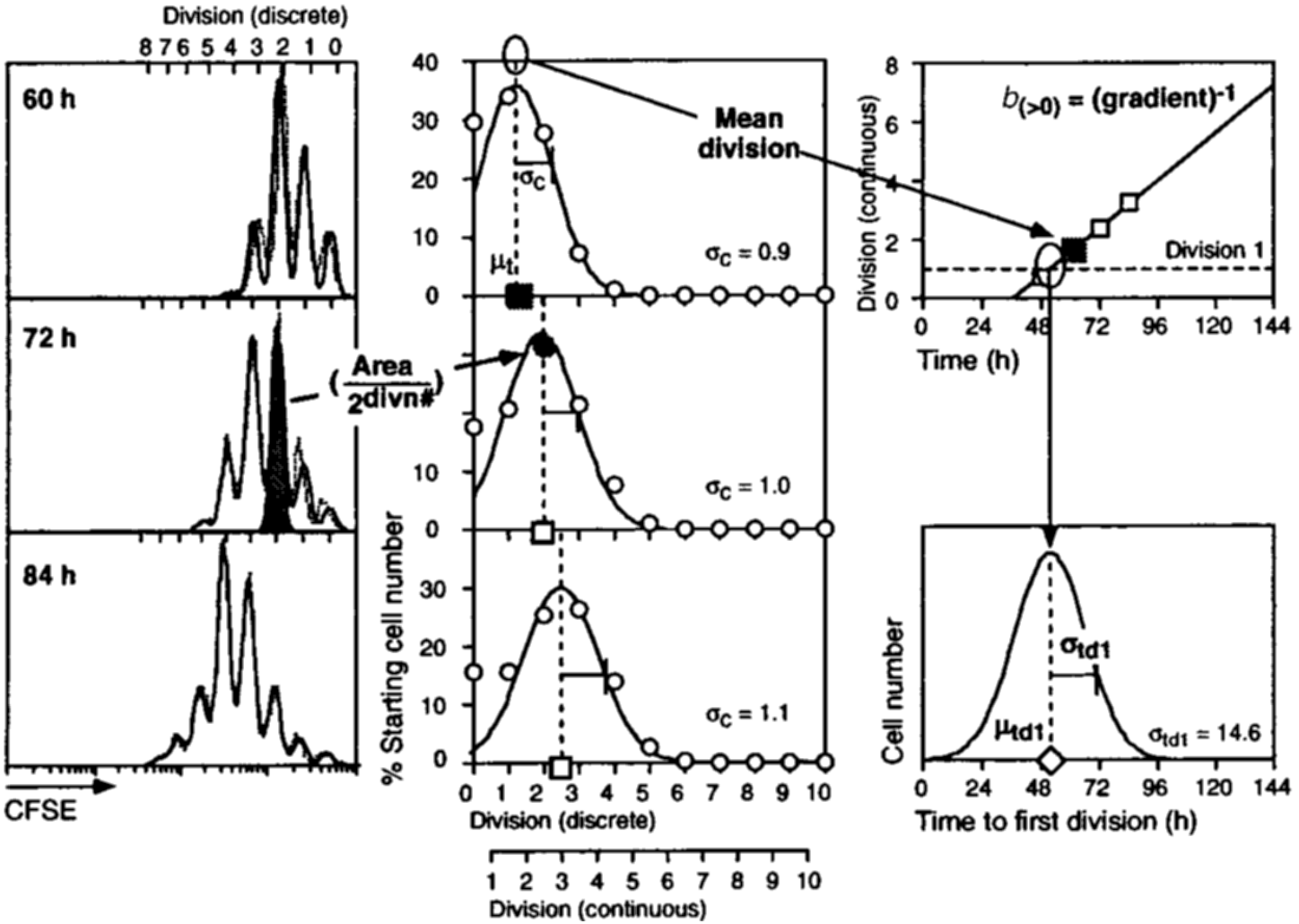
- Label cells *in vitro*: cell pick up the dye
- Two daughter cells after division half the intensity
- Follow cells *in vitro* or *in vivo* for 7 divisions
- Because method tracks individual divisions it is typically more informative than BrdU or ^{2}H-glucose labeling

Example



After 48h **most** cells have completed three divisions:
Not true because division index 3 naturally has 2^3-fold more cells

Gett & Hodgkin, Nature Immunology, 2000



Gett & Hodgkin, Nature Immunology, 2000

- Data are fingered: fit a log-normal Gaussian distribution

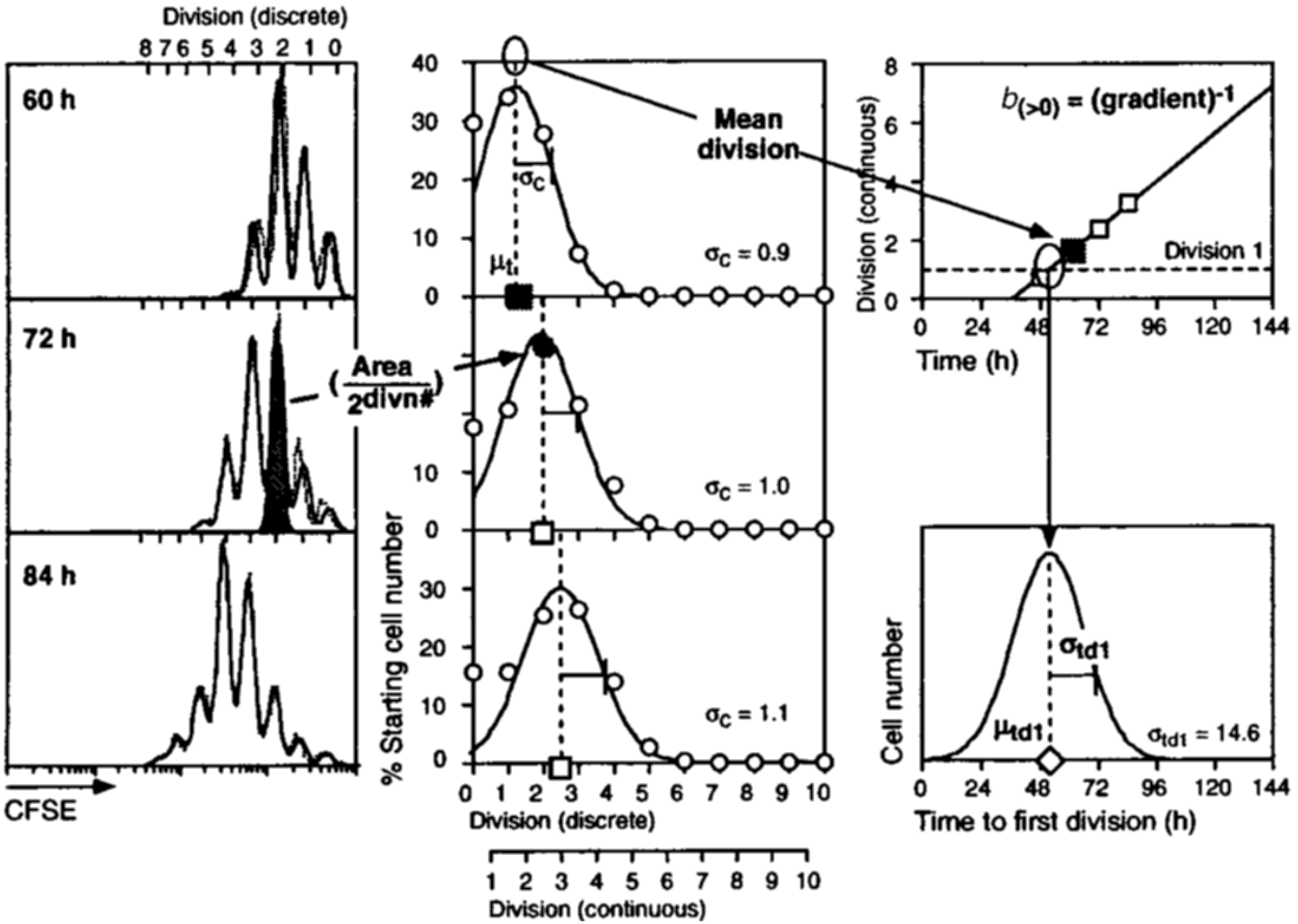
$\rightarrow$ number (or fraction) of cells in each division index i

- Divide this by 2^i to correct for number of divisions

$\rightarrow$ otherwise **overestimation** of highest division index

- Precursor cohort plot: frequency distribution of normalized cell numbers (or fractions)
- Compute mean of frequency distribution for divided cells $\widehat{\mu_2}(t) = \sum_{i=1} \infty i f_i(t)$
- Conjecture: mean $\widehat{\mu_2}(t)$ increases linearly in time!

Gett & Hodgkin, Nature Immunology, 2000

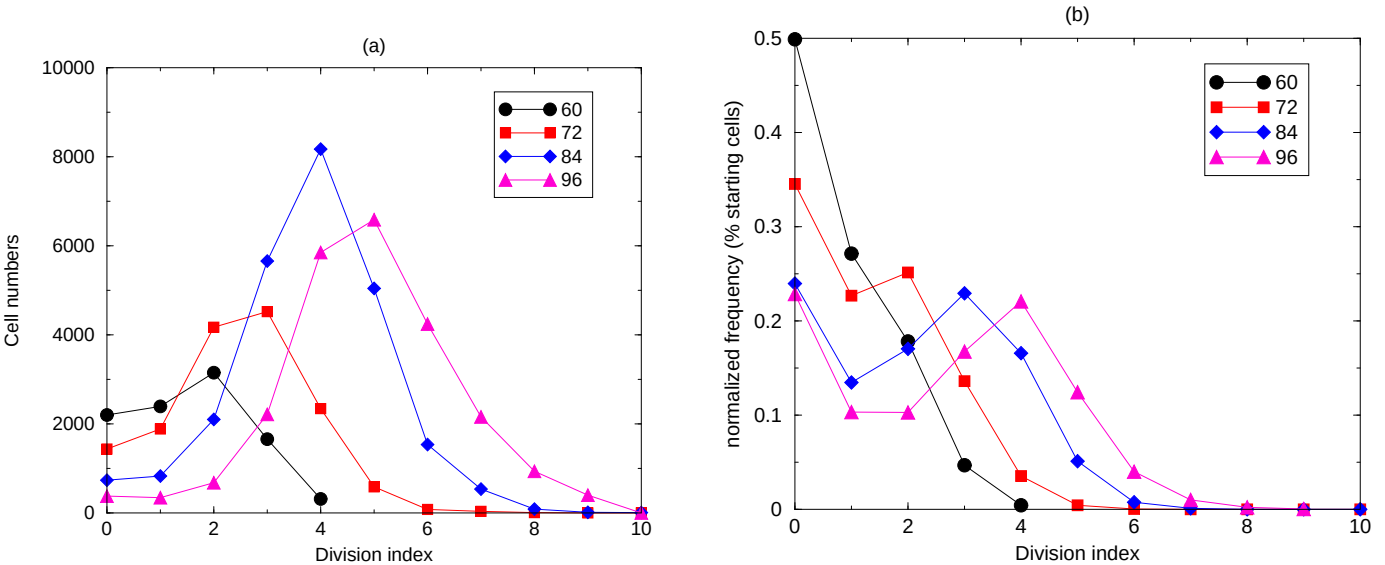


Gett & Hodgkin, Nature Immunology, 2000

Mean increases linearly in time:

- Slope reflects division time
- Time at which $\widehat{\mu_2} = 1$ is time to first division
- Cell cycle times of 20 h and 60 h.
- Intuitive leap: if true this seems a very general approach
- Frequency distributions seem Gaussian: times to first divisions have a Gaussian distribution.

Normalization gives fractions completing n divisions



Importantly, the 2^n normalization repairs the mistake of saying that **most** cells have completed n divisions

Questions

- Pilyugin et al. JTB (in press) showed for homogeneous models that slope of mean depends the distribution of the death rates.
- Here we focuss on heterogeneous model: resting cells that are stimulated to divide.
- Do Gaussian frequency distributions truly reflect a Gaussian distribution in the time to first division?

Modeling the Gett & Hodgkin approach

Homogeneous case:

$$\begin{aligned} \frac{\mathrm{d}N_0}{\mathrm{d}t} &= -(p+d)N_0 \\ \frac{\mathrm{d}N_i}{\mathrm{d}t} &= 2pN_{i-1} - (p+d)N_i \ , \qquad \text{for } i = 1, \ldots, \infty \ . \end{aligned}$$

The total number of cells: $N(t) = N_0(0)\mathrm{e}^{(p-d)t}$.

The frequency distribution of cells over the division numbers, is a Poisson distribution: $\mu(t) = 2pt$,

$$F_i(t) = \frac{(2pt)^i}{i!}\mathrm{e}^{-2pt} \ , \qquad \text{for } i = 1, \ldots, \infty$$

Normalization

When $n_i(t) \equiv N_i(t)/2^i$ one obtains

$$f_i(t) = \frac{(pt)^i}{i!}\mathrm{e}^{-pt} \ ,$$

and

$$n(t) = N_0(0)\mathrm{e}^{-dt} \ .$$

Thus $\mu_2(t) = pt$ and $\mu_2 = 1$ yields $t = 1/p$.

Death rate can be estimated from $n(t) = N_0(0)\mathrm{e}^{-dt}$.

Mean of divided cells

$$\widehat{\mu_2}(t) \equiv \sum_{i=1}^{\infty} i f_i(t) / \sum_{i=1}^{\infty} f_i$$

This new mean is

$$\widehat{\mu_2}(t) = \frac{pt}{1 - \mathrm{e}^{-pt}} \ ,$$

and the normalized number of dividing cells is

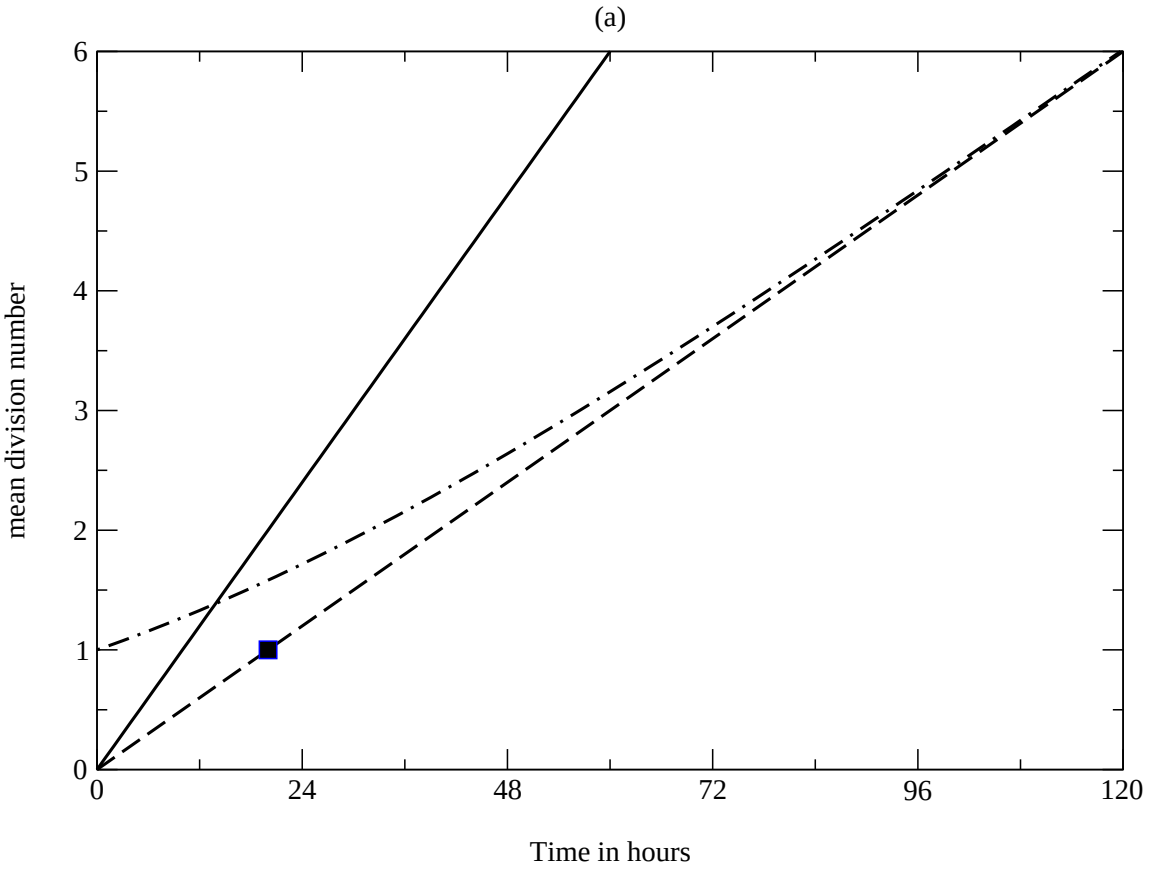
$$\widehat{n}(t) = N_0(0)\mathrm{e}^{-dt}\left[1 - \mathrm{e}^{-pt}\right] \ .$$

Both have an initial transient of one cell cycle, p^{-1}.

Moreover, solving $\widehat{\mu_2}(t) = 1$ gives zero ($\widehat{\mu_2}(0) \to 1$).

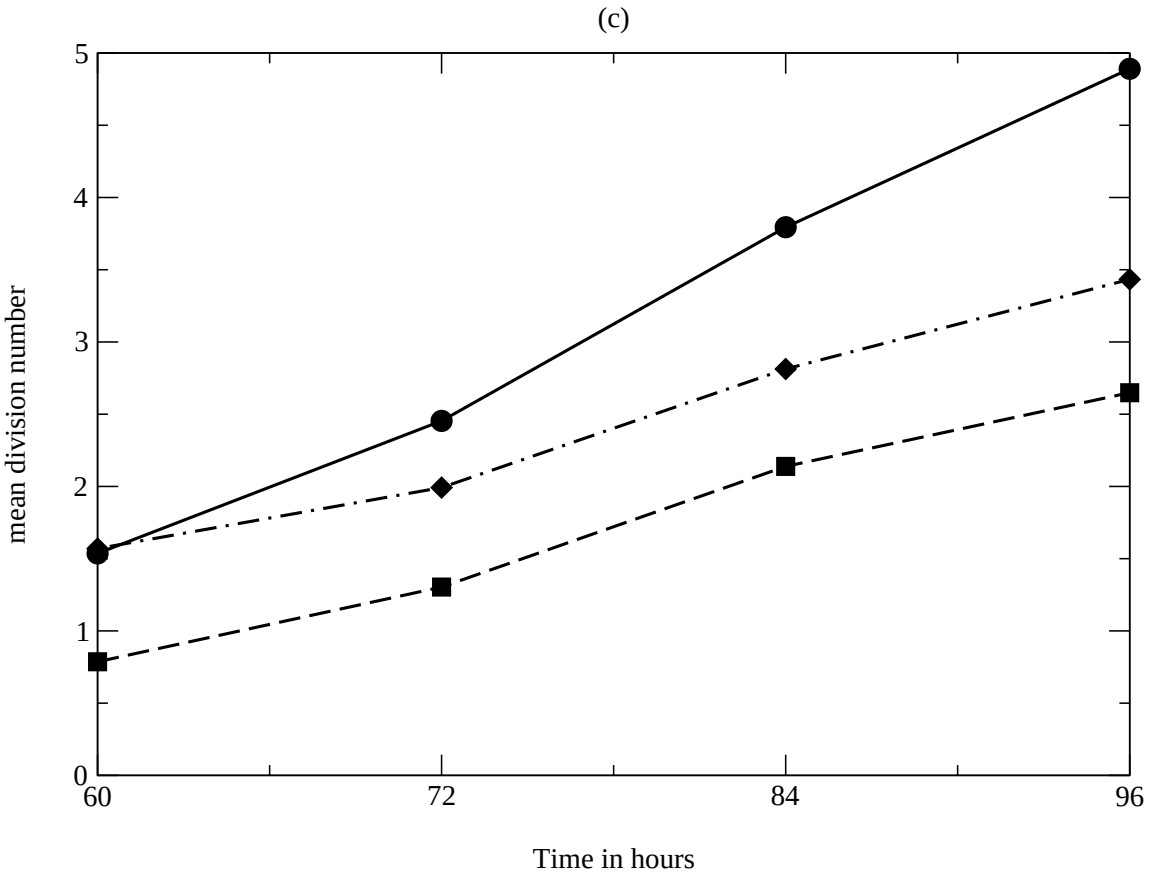
Only after this transient $\widehat{\mu_2}(t) \to pt$.

Homogeneous model: three means



Use μ_2 or asymptote of $\widehat{\mu_2}$ to estimate p.

Means of the data



Increase seems fairly linear, asymptotic regime approached?
Why use $\widehat{\mu_2}(t)$ instead of $\mu_2(t)$ or even simply $\mu(t)$?

Conclusion

Method	Transient		cells	Intersect	Slope
μ	0	0	$\mathrm{e}^{(p-d)t}$	$(2p)^{-1}$	$2pt$
μ_2	0	0	e^{-dt}	p^{-1}	pt
$\widehat{\mu_2}$	p^{-1}	20 h	e^{-dt}	p^{-1}	pt

Fraction of cells that never divides

ϕ is fraction of precursors cells that divides, and let τ be the time delay before proliferation starts.

The gives the total normalized cell numbers

$$n(t) = P\mathrm{e}^{-d(t+\tau)}$$

and the frequency distribution

$$f_0(t) = \phi\mathrm{e}^{-pt} + 1 - \phi \quad \text{and} \quad f_i(t) = \phi\,\frac{(pt)^i}{i!}\mathrm{e}^{-pt}\ ,$$

with mean

$$\mu_2(t) \equiv \sum_{i=0}^{\infty} i f_i(t) = \phi pt\ .$$

Solving $\mu_2(t) = 1$ also fails to deliver the time to first division.

The Gett & Hodgkin mean $\widehat{\mu_2}(t)$

Because ϕ cancels from when one computes

$$\widehat{\mu_2}(t) \equiv \sum_{i=1}^{\infty} i f_i(t) / \sum_{i=1}^{\infty} f_i \ ,$$

one obtains the same mean as before:

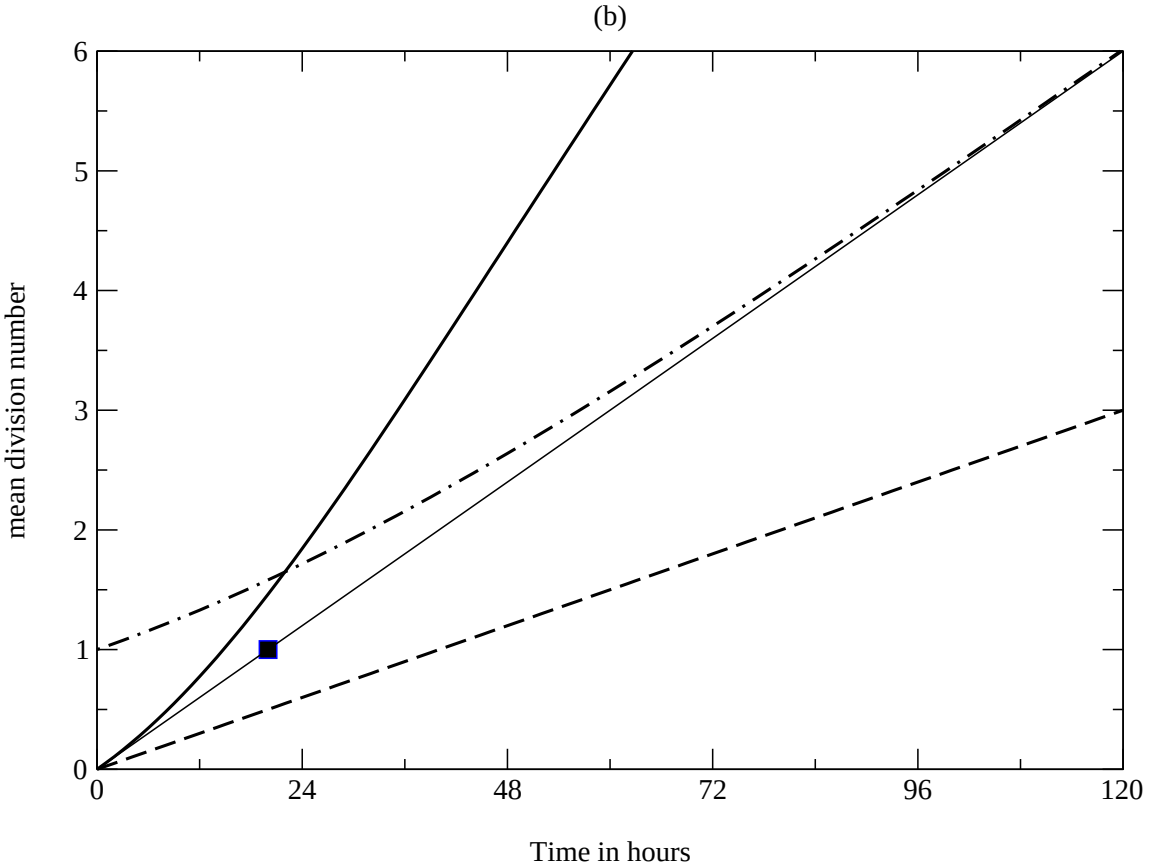
$$\widehat{\mu_2}(t) = \frac{pt}{1 - \mathrm{e}^{-pt}} \ ,$$

and the normalized number of dividing cells is

$$\widehat{n}(t) = N_0(0)\mathrm{e}^{-dt}\left[1 - \mathrm{e}^{-pt}\right] \ .$$

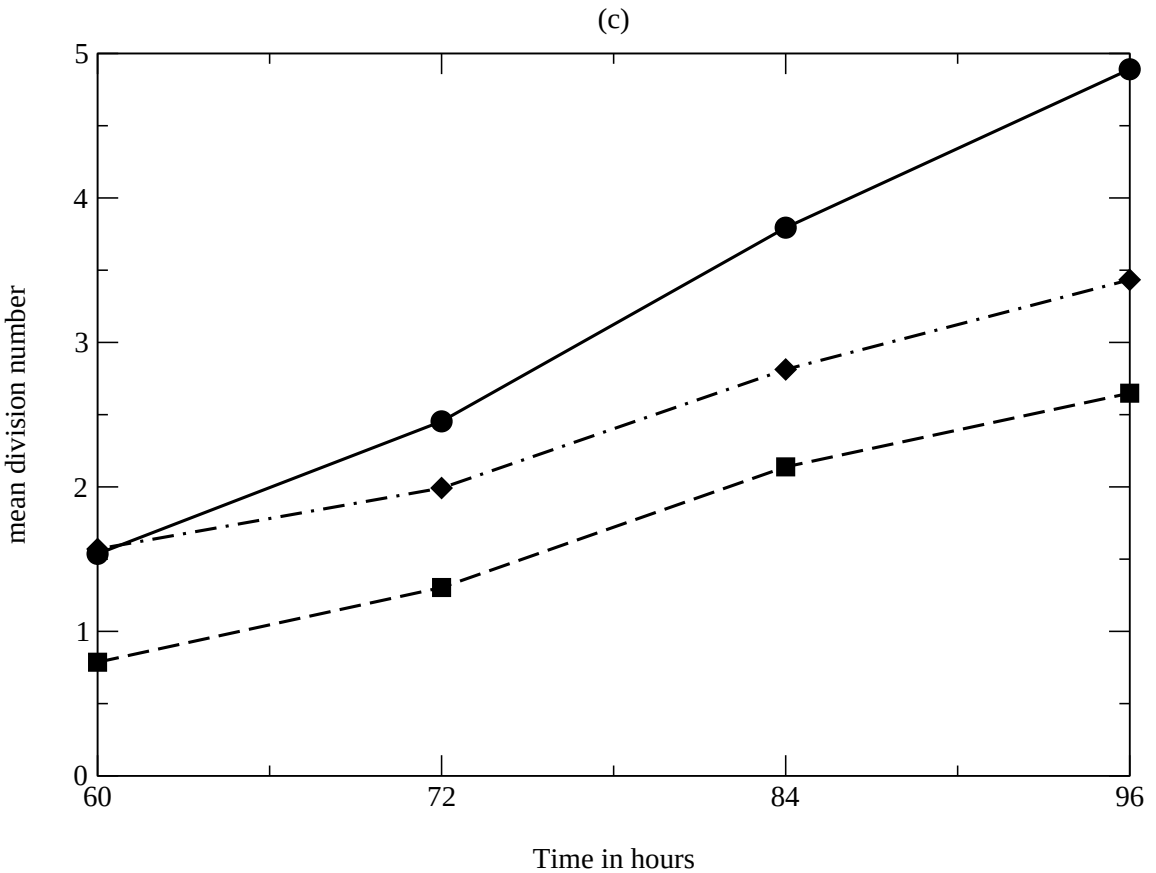
Thus after the initial transient of one cell cycle, p^{-1}, one should be able to estimate p.

Fraction of cells that never divides: $\phi = 0.5$



$\widehat{\mu_2}(t)$ seems to perform better than $\mu_2(t)$.

Means of the data

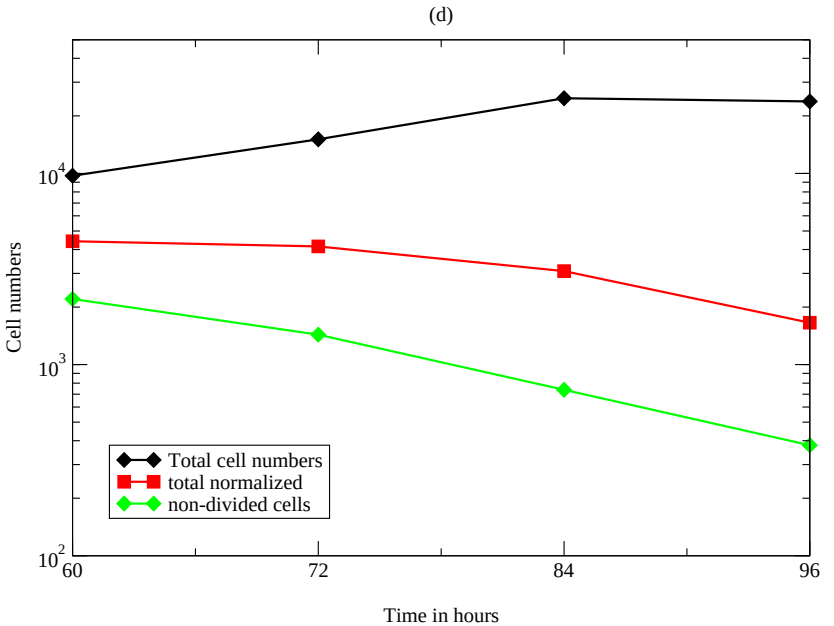


Data seem to suggest that $\phi = 1$.

Conclusion

Method	Transient		cells	Intersect	Slope
μ	0	0	$e^{(p-d)t}$	$(2p)^{-1}$	$2pt$
μ_2	0	0	e^{-dt}	p^{-1}	pt
$\widehat{\mu_2}$	p^{-1}	20 h	e^{-dt}	p^{-1}	pt
μ_2	τ	40 h	e^{-dt}	$\tau + (\phi p)^{-1}$	ϕpt
$\widehat{\mu_2}$	$\tau + p^{-1}$	60 h	e^{-dt}	$\tau + p^{-1}$	pt

Heterogeneous case



Slopes of 0.025h^{-1}, -0.025h^{-1}, and -0.05h^{-1}. Death rate $d = 0.025\text{h}^{-1}$, delivers $p = 0.05\text{h}^{-1}$. Loss of non-divided cells would be $N_0(t) = N(0)\text{e}^{-(p+d)t} = N(0)\text{e}^{-0.05t}$.

No evidence for $\phi < 1$

Heterogeneous model

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = -(p'+d')N_0 \quad , \qquad \frac{\mathrm{d}N_1}{\mathrm{d}t} = 2p'N_0 - (p+d)N_1$$
$$\frac{\mathrm{d}N_i}{\mathrm{d}t} = 2pN_{i-1} - (p+d)N_i \ , \qquad \text{for } i = 2, \ldots, \infty \ ,$$

Total cell numbers obey

$$N(t) = \frac{N_0(0)\mathrm{e}^{(p-d)t}}{c}\left[2p' + b\mathrm{e}^{-ct}\right] \ ,$$

and the mean is

$$\mu(t) = \frac{2p'[a(\mathrm{e}^{-ct}-1) + 2pct]}{c[2p' + b\mathrm{e}^{-ct}]}$$

where $a = p - p' + d - d' \geq 0$, $b = p - p' - (d - d') > 0$, and a transient of $c = p + p' - (d - d') > 0$.

Heterogeneous model: $\mu(t)$

For times larger than $1/c = [p + p' - (d - d')]^{-1}$ h the mean will approach

$$\mu(t) = 2pt - \frac{p - p' + d - d'}{p + p' - (d - d')} \ ,$$

which increases with the expected slope $2pt$.

Solving $\mu(t) = 1$ from this asymptote gives $t = 1/(p+p'-d+d')$, which only delivers the time to first division when $d = d'$.

Picking $p = 0.05\text{h}^{-1}$, $p' = 1/60\text{h}^{-1}$, $d = 0.025\text{h}^{-1}$, and $d' = 0.01\text{h}^{-1}$, the transient is about 20h.

Heterogeneous model: $\mu_2(t)$

Similar analysis gives

$$\mu_2(t) = \frac{p'}{\gamma} \frac{\gamma p t + a(\mathrm{e}^{-\gamma t} - 1)}{p' + (d' - d)\mathrm{e}^{-\gamma t}}$$

and

$$n(t) = \frac{N_0(0)\mathrm{e}^{-dt}}{\gamma} \left[p' + (d' - d)\mathrm{e}^{-\gamma t}\right] ,$$

which both have a transient of $\gamma = p' + d' - d = p - a$.

After this transient, i.e., for $t \to \infty$

$$\mu_2(t) = pt - \frac{a}{\gamma} = pt - \frac{p - p' + d - d'}{p' + d' - d} ,$$

which increases with slope pt.

But ...

Solving $\mu_2(\infty) = 1$ yields $t = 1/(p' + d' - d)$ which is only equal to the time to first division when $d = d'$.

Picking $p = 0.05\text{h}^{-1}$, $p' = 1/60\text{h}^{-1}$, $d = 0.025\text{h}^{-1}$, and $d' = 0.01\text{h}^{-1}$, the transient is about $1/\gamma = 600\text{h}$.

Heterogeneous model: $\widehat{\mu_2}(t)$

Similar analysis yields

$$\widehat{n}(t) = \frac{p' N_0(0) \mathrm{e}^{-dt}}{\gamma} \left[1 - \mathrm{e}^{-\gamma t}\right] ,$$

and

$$\widehat{\mu_2}(t) = \frac{\gamma p t + a(\mathrm{e}^{-\gamma t} - 1)}{\gamma[1 - \mathrm{e}^{-\gamma t}]} .$$

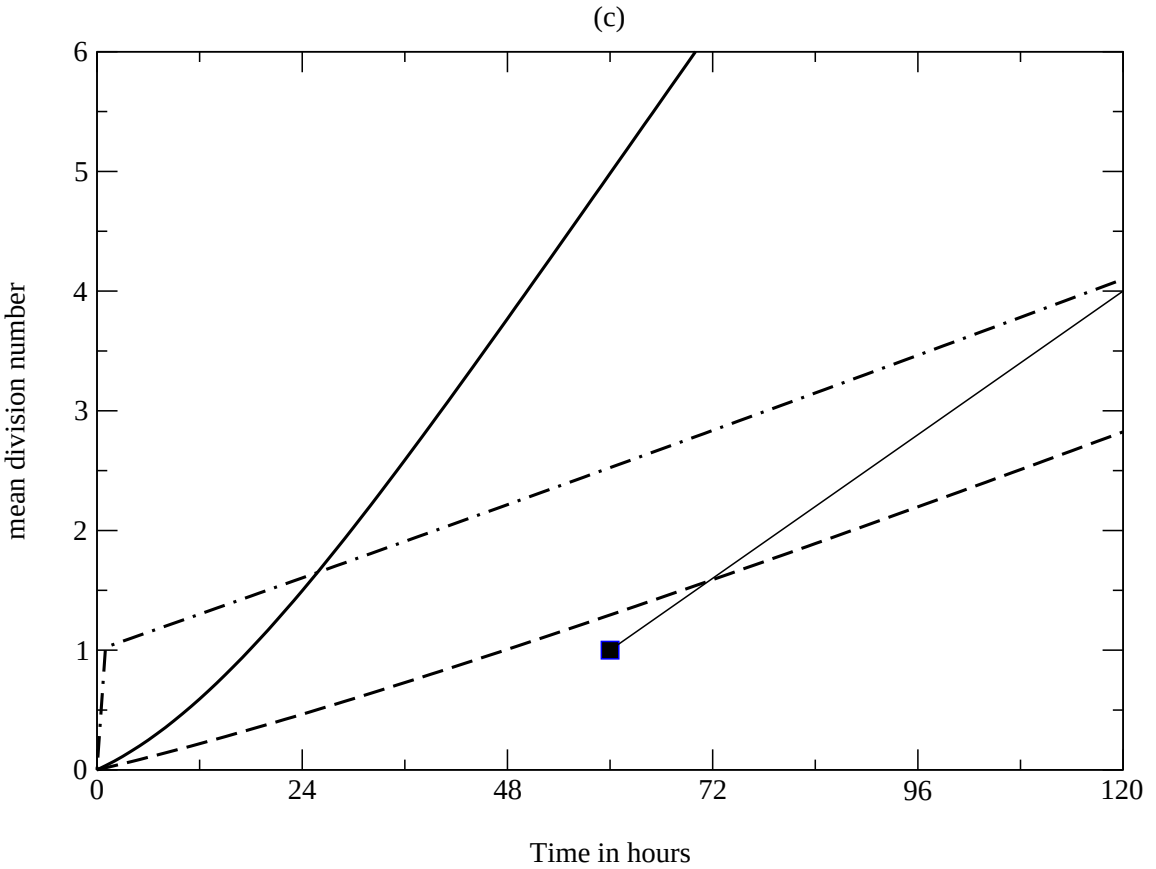
with the same long transient γ.

For $t \to \infty$ the mean of the divided cells approaches

$$\widehat{\mu_2}(t) = pt - \frac{a}{\gamma} = pt - \frac{p - p' + d - d'}{p' + d' - d}$$

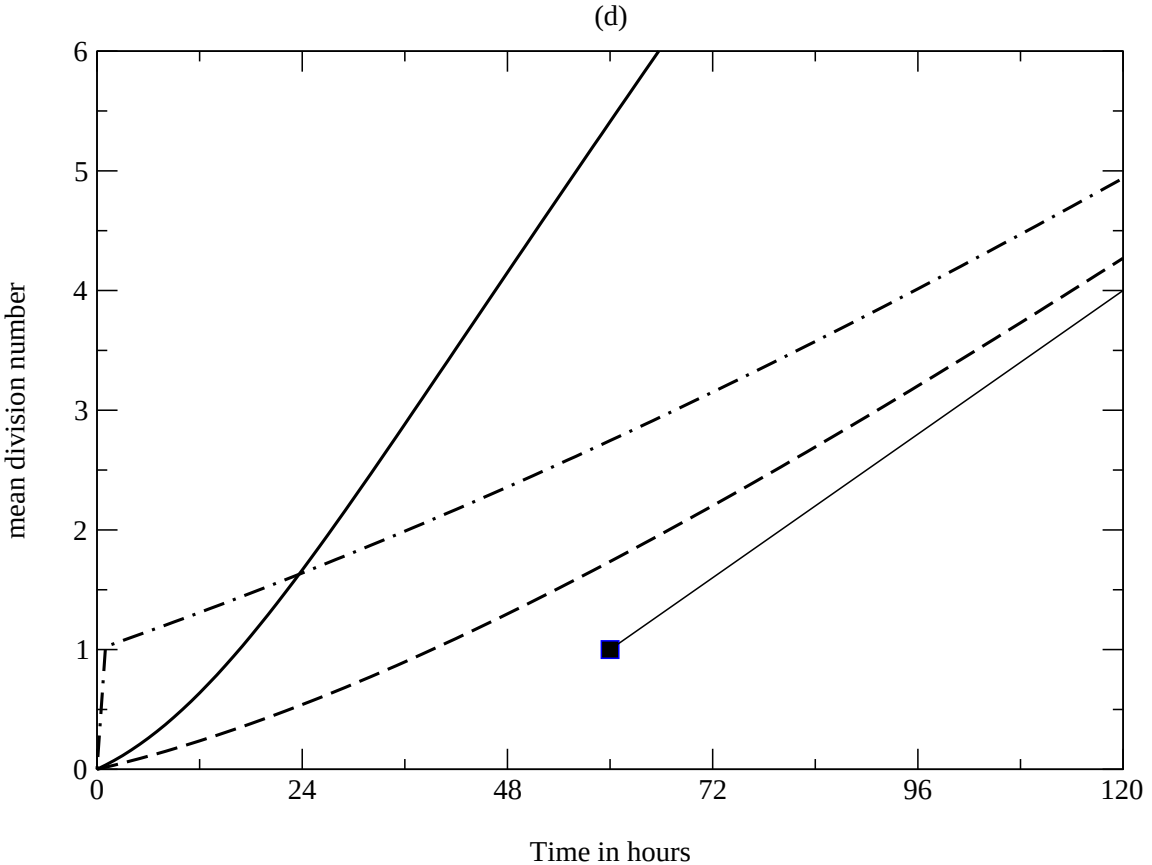
which is the same as $\mu_2(t)$.

Heterogeneous model: three means



Note that γ could even be negative.

Heterogeneous model: three means for $d' = d$

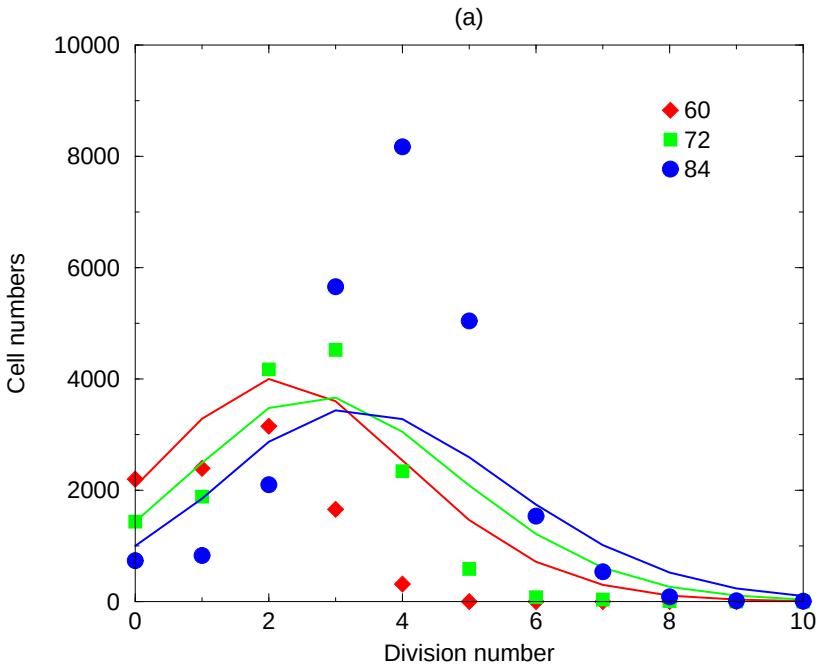


Difference between death rates determines length of the transient.

Conclusion

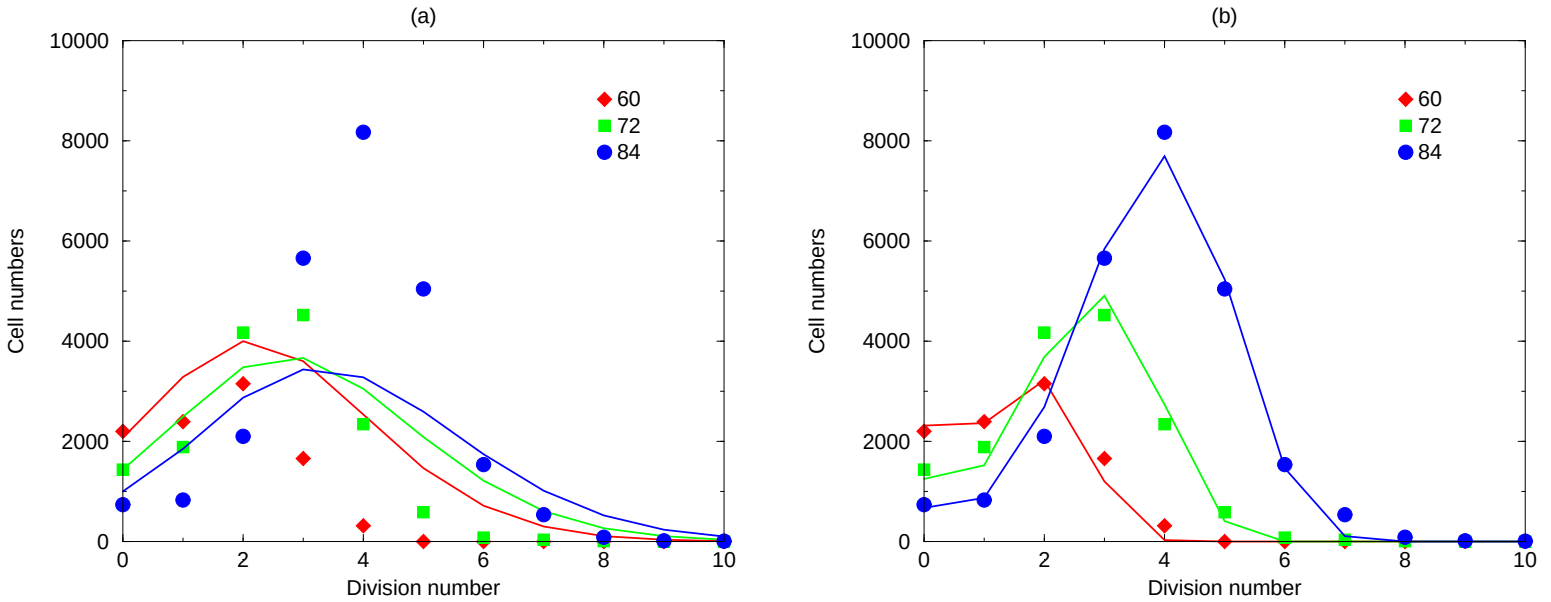
Method	Transient		cells	Intersect	Slope
μ	0	0	$e^{(p-d)t}$	$(2p)^{-1}$	$2pt$
μ_2	0	0	e^{-dt}	p^{-1}	pt
$\widehat{\mu_2}$	p^{-1}	20 h	e^{-dt}	p^{-1}	pt
μ_2	τ	40 h	e^{-dt}	$\tau + (\phi p)^{-1}$	ϕpt
$\widehat{\mu_2}$	$\tau + p^{-1}$	60 h	e^{-dt}	$\tau + p^{-1}$	pt
μ	$[p + p' - (d - d')]^{-1}$	19.35 h	$e^{(p-d)t}$	$[p + p' - d + d']^{-1}$	$2pt$
μ_2	$[p' + d' - d]^{-1}$	600 h	e^{-dt}	$[p' + d' - d]^{-1}$	pt
$\widehat{\mu_2}$	$[p' + d' - d]^{-1}$	600 h	e^{-dt}	$[p' + d' - d]^{-1}$	pt

Fitting with the ODE model



Estimates: $p = 0.025\text{h}^{-1}$ (40h), $p' = 0.022\text{h}^{-1}$ (45)h, $d' = 0.01\ \text{h}^{-1}$, and $N(0) = 1.5 \times 10^4$ cells.

Fitting with the Smith-Martin model



ODE models perform poorly: time delay is required.

Conclusions

- Collect data **late enough** to approach the linear regime of $\mu(t)$, but **early enough** to exclude confounding factors
- Parameter estimation using means sensitive to transients
- Since $\mu(t)$ has the shortest transient one could argue that this mean is the most reliable?
- Normalization remains important to argue about fractions of cells having completed n divisions, and to test whether means are increasing linearly.
- Difference between $\mu_2(t)$ and $\widehat{\mu_2}(t)$ may give indication that a fraction of the cells fails to divide
- Time to first division very difficult to estimate
- Plot $N(t)$, $n(t)$, and $N_0(t)$ and estimates slopes.
- Use all this information as an initial guess for fitting with the Smith-Martin model.
- **Do not fit ODE models to CFSE data**