

# SCIENTIFIC REPORTS



OPEN

## Estimating dose-specific cell division and apoptosis rates from chemo-sensitivity experiments

Yiyi Liu<sup>1</sup> & Forrest W. Crawford<sup>1,2,3</sup>

*In-vitro* chemo-sensitivity experiments are an essential step in the early stages of cancer therapy development, but existing data analysis methods suffer from problems with fitting, do not permit assessment of uncertainty, and can give misleading estimates of cell growth inhibition. We present an approach (bdChemo) based on a mechanistic model of cell division and death that permits rigorous statistical analyses of chemo-sensitivity experiment data by simultaneous estimation of cell division and apoptosis rates as functions of dose, without making strong assumptions about the shape of the dose-response curve. We demonstrate the utility of this method using a large-scale NCI-DREAM challenge dataset. We developed an R package “bdChemo” implementing this method, available at <https://github.com/YiyiLiu1/bdChemo>.

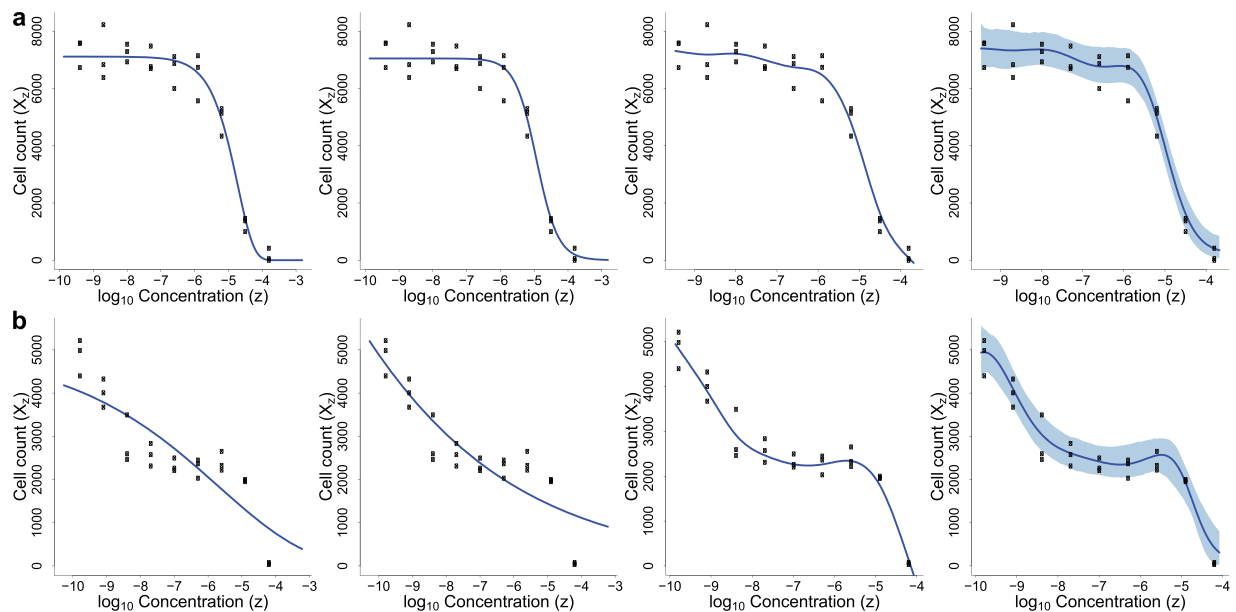
In the early stages of cancer therapy development, potencies of candidate compounds are usually tested *in vitro* through chemo-sensitivity studies<sup>1,2</sup>. Researchers treat cultured tumor cells with different concentrations of compounds, and the numbers of cells remaining after a follow-up time  $T$  are recorded via fluorescent signal intensities that measure general metabolism levels or enzymatic activities<sup>3</sup>. Compounds that achieve desired tumor inhibition effects within dose ranges that are not considered clinically toxic are identified for further optimization and then tested with animal models and clinical trials<sup>1</sup>. Given the significant investment required to bring drug candidates to preclinical and clinical stages<sup>4</sup>, screening and selecting the most promising candidates from chemo-sensitivity studies is essential for drug development.

Conventionally, the growth inhibition response of a cell line to a compound is modeled with a sigmoid curve. The most commonly used are the Gompertz<sup>5,6</sup> and logistic curves<sup>7,8</sup>. Figure 1a (left and middle panels) illustrates a typical dataset (cell line AU565 treated with compound 4-HC<sup>6</sup>), with fitted Gompertz and logistic curves. Concentrations needed to achieve certain levels of inhibition effects, such as  $GI_{50}$  (growth of the cell population is inhibited by half),  $TGI$  (growth of the cell population is eliminated) and  $LC_{50}$  (half of the initial cell population is eliminated) are then estimated as assessments of the compound's potency<sup>9</sup>.

However, this approach suffers from problems that may hinder its utility in chemo-sensitivity evaluation. First, it relies on a parametric form of the growth curve. Different compounds may affect cell growth in physiologically distinct ways makes it unreasonable to believe that all inhibition patterns, which result from complex interactions between compounds and cells, could be modeled with the same parametric form (Fig. 1b, the first two panels). A newer nonparametric method called grofit<sup>10</sup> provides a framework for fitting more flexible dose-response curves using spline smoothing. However, all these methods, by fitting a single growth curve, only deliver information about the combined effects of the compound on cell birth and death processes, while understanding these responses individually is critical to designing experiments that more elaborately investigate the compound's mechanism of action<sup>11</sup>. Moreover, cancer therapy development often begins by designing compounds that target pathways either inhibiting cell division or inducing cell death separately<sup>12</sup>, so it would be helpful to separately discern these effects from early-stage chemo-sensitivity experiments. Finally, existing approaches consider only “point estimation” of  $GI_{50}/TGI/LC_{50}$ , as a summary of the compound inhibition effects without providing measures of uncertainty for these estimated quantities.

To overcome these limitations, we describe a new analysis approach (bdChemo) for chemo-sensitivity studies. This method specifies a mechanistic model of stochastic cell growth, fits semi-parametric dose-response curves without strong assumptions on their functional forms, and separately estimates dose-specific “birth” and “death” rates for the compound. For a given compound, we assume that a cell line's per-cell birth and death rates,  $\lambda$  and

<sup>1</sup>Department of Biostatistics, Yale School of Public Health, New Haven, USA. <sup>2</sup>Department of Ecology and Evolutionary Biology, Yale University, New Haven, USA. <sup>3</sup>Yale School of Management, New Haven, USA. Correspondence and requests for materials should be addressed to F.W.C. (email: [forrest.crawford@yale.edu](mailto:forrest.crawford@yale.edu))



**Figure 1.** Dose-response modeling of cell line AU565 treated with compound 4-HC (**a**) and cell line SUM52PE treated with compound Everolimus (**b**), using Gompertz curve fitting (the first panel), logistic curve fitting (the second panel), grofit (the third panel) and bdChemo (the fourth panel). Solid circles and blue curves are data points and fitted curves, respectively. For bdChemo, curves and shades represent the posterior mean and the 95% credible interval of the Kendall Process mean, respectively.

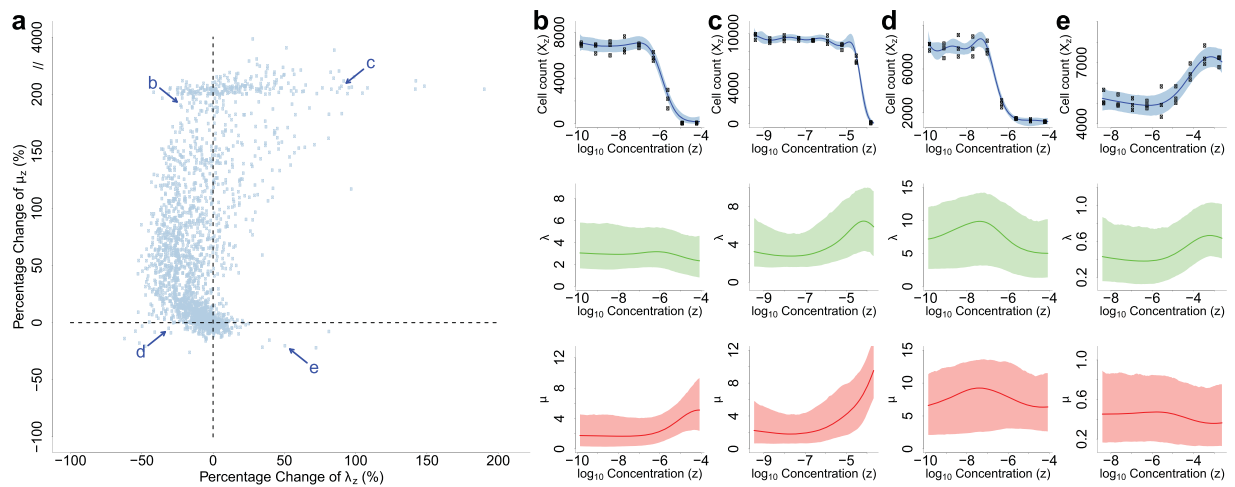
$\mu$ , are time-homogenous and depend on the  $\log_{10}$  concentration  $z$ , of the compound; a cell community with size  $k$  has aggregate population-level birth and death rates  $k\lambda(z)$  and  $k\mu(z)$ , respectively. Such a system is known as Kendall, or birth-death, process<sup>13</sup> (BDP). When the BDP accurately characterizes of the dynamics of cellular response to the compound, the estimated rates of birth and death may have a mechanistic interpretation as “cell division” and “apoptosis” rates, respectively. To avoid unnecessary assumptions on the dose-dependent shapes of  $\lambda(z)$  and  $\mu(z)$  and allow a flexible relationship between dose and response, we employ a semi-parametric Bayesian approach by assigning Gaussian process<sup>14</sup> priors, which treats a regression function on a continuous domain as an infinite-dimensional random variable. The method assumes that any finite marginal distribution follows a multivariate Gaussian distribution without restrictions on the parametric form of the regression function. We estimate the dose-response relationships of the per-cell birth and death rates  $\lambda(z)$  and  $\mu(z)$  as well as other model parameters using a Markov Chain Monte Carlo (MCMC) algorithm<sup>15</sup>. Uncertainty in estimates of birth and death rates is appropriately propagated into uncertainty in summary statistics like  $GI_{50}$ ,  $TGI$  and  $LC_{50}$ .

We demonstrate the utility of this method on a large-scale chemosensitivity dataset from NCI-DREAM challenge containing cell population size measurements on 53 breast cancer cell lines treated by 28 compounds. In the original work, the authors fit a Gompertz curve<sup>5</sup> for each experiment and calculated  $GI_{50}$  to quantify the sensitivity of the cell line to the compound. We apply bdChemo to the dataset and estimate the posterior mean as well as the 95% equal quantile credible intervals (CI) of chemosensitivity summary statistics  $GI_{50}$ ,  $TGI$  and  $LC_{50}$ , and cell birth and death rates,  $\lambda(z)$  and  $\mu(z)$ , for each compound and cell line combination. We compare the results of the proposed method with those obtained by conventional Gompertz and logistic curve fitting approaches as well as grofit.

## Results

We analyze the data from NCI-DREAM drug sensitivity prediction challenge<sup>6</sup> to demonstrate the utility of estimates produced by the proposed method. This dataset contains dose-response measurements of 28 compounds on 53 breast cancer cell lines. Cells were treated with 9 doses of each compound in triplicate and cell counts at 72 h post treatment were measured using the Cell Titer Glo assay. In the original work, the authors fit a Gompertz curve<sup>5</sup> for each experiment (a cell line treated by a compound) and calculated  $GI_{50}$  (a point estimate without uncertainty evaluation) to quantify the sensitivity of the cell line to the compound. We summarize the posterior mean and 95% equal quantile credible intervals (CI) of  $\lambda(z)$ ,  $\mu(z)$ ,  $GI_{50}$ ,  $TGI$  and  $LC_{50}$  returned by bdChemo in Supplementary Table S1 and Figure S1.

**bdChemo fits dose-response curve flexibly.** Conventional Gompertz and logistic curve fitting approaches rely on specified parametric forms of the dose-response curve. These parametric forms may not be flexible enough to describe the observed dose-response data due to the complex interactions between compounds and cell lines. Compared to the restricted parametric curve fitting approaches, bdChemo does not put strong assumptions on the functional forms of the dose-response curve, and hence provides a flexible and data-driven approach to study the effect of a compound on a cell line.



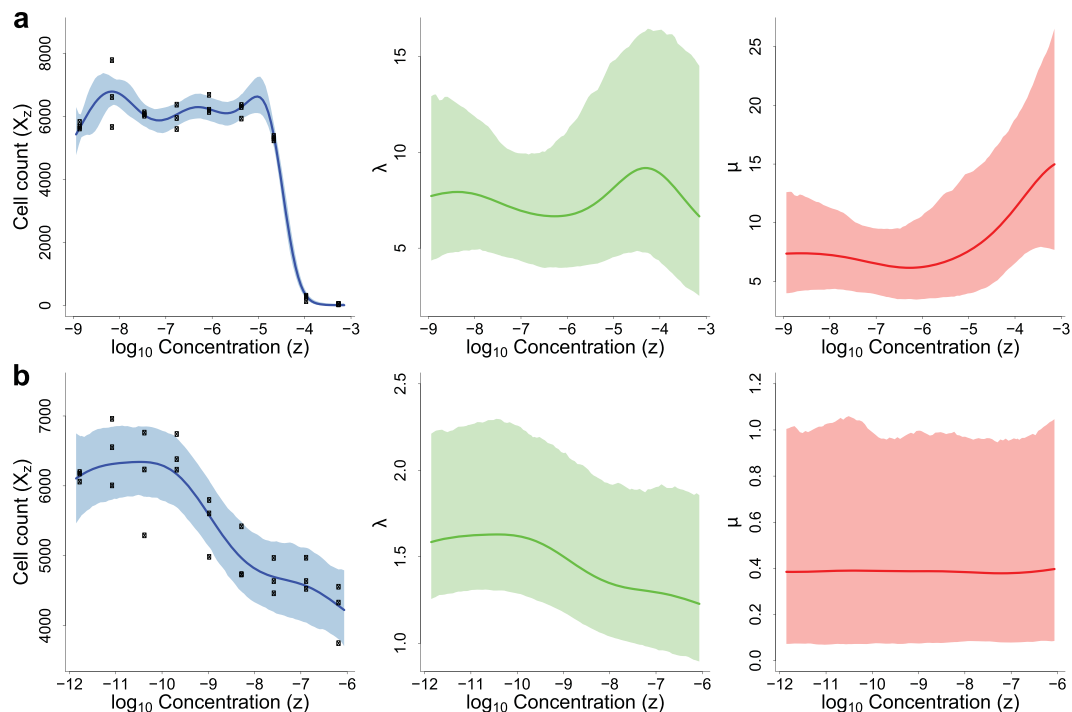
**Figure 2.** (a) Percentage change between the posterior mean estimates of death rates ( $\mu(z)$ ) at the largest and the smallest tested concentrations v.s percentage change between posterior means of birth rates ( $\lambda(z)$ ). (b–e) Posterior estimations of Kendall Process means (dose-response curves, upper panel), birth rates (middle panel) and death rates (lower panel) of cell line HCC1954 treated with compound IKK 16 (b), cell line MCF7 treated with compound 4-HC (c), cell line HCC1428 treated with compound MG-132 (d) and cell line HCC202 treated with compound PS-1145 (e). Curves and shades represent the posterior mean and the 95% credible interval of the corresponding variables, respectively.

Figure 1a and b depict two examples of different growth patterns modeled by the four approaches. For compound 4-HC working on cell line AU565 (Fig. 1a), the growth curve follows a sigmoid shape. In such case, all four approaches produced good fits. However, for compound Everolimus working on cell line SUM52PE (Fig. 1b), where there is a plateau around the waist of the growth curve, the two sigmoid cell fitting approaches (the first two panels) could not capture this trend, while growth (the third panel) and bdChemo (the fourth panel) generated curves more concordant with the data.

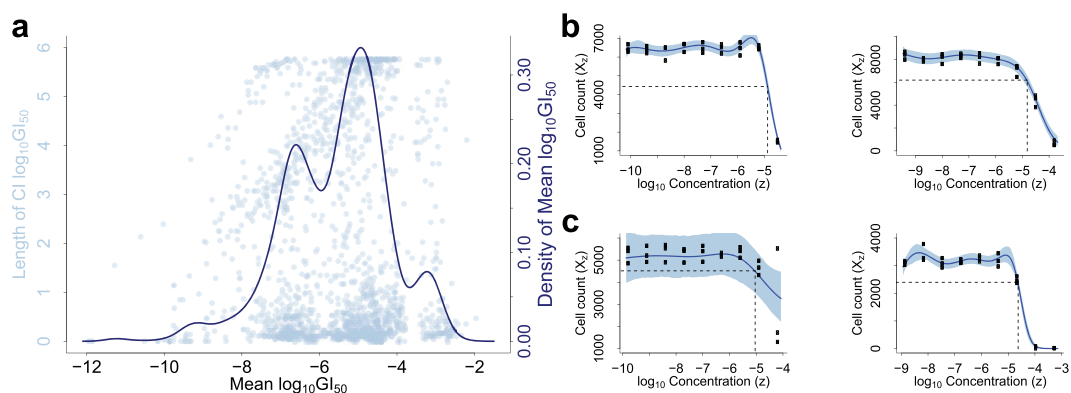
**bdChemo provides separate estimates of cell birth and death rates.** Together, cell division and apoptosis, as functions of compound dose, determine the dynamics of cell line response. For these distinct cellular processes, model-based estimation of compounds' effects on cell birth and death rates can serve as a screening tool for candidate compounds and provide guidance for hypothesis generation and experiment design to study compounds' cellular mechanisms of action. We depict the percentage changes between the posterior means of birth/death rates at the largest and the smallest tested concentrations in Figure 2a (all compounds) and Figure S2 (individual compound). The overall growth, birth and death curves of four examples are included in Figure 2b–e as further illustrations. When treated by a compound, most cell lines in this study show decreased birth and increased death rates under higher concentrations (top left part of Figure 2a; an example in Figure 2b), while a few exhibit only decreased birth rates or increased death rates alone (top right and bottom left parts in Figure 2a; examples in Figure 2c,d). Not all compounds in this study appear to inhibit cell growth, resulting in a few other points with large birth rate increment or death rate decrement effect (bottom right part in Figure 2a; an example in Figure 2e).

These scenarios may suggest different underlying mechanisms of action. For example, in several other cancers, Imatinib has been reported to kill tumor cells by decreasing the activity of tyrosine kinase enzymes<sup>16</sup>, whereas Cetuximab was known to hinder uncontrolled tumor cell division as an EGFR inhibitor<sup>17</sup>; here we observe similar effects on breast cancer cell lines: for the compound Imatinib and cell line BT474 (Figure 3a), birth rate  $\lambda(z)$  is relatively stable with respect to dose  $z$  in the tested range, while death rate  $\mu(z)$  first stays steady but then increases rapidly when dose  $z$  becomes large; in contrast, for Cetuximab working on cell line HCC1806 (Figure 3b), birth rate  $\lambda(z)$  decreases while death rate  $\mu(z)$  stays stable. Therefore, we may hypothesize that compound Imatinib mainly works by inducing cell apoptosis on BT474 while compound Cetuximab is more likely to target on blocking cell division on HCC1806, and design experiments to investigate the effects of these compounds on cell apoptosis and division-related pathways, respectively, to better understand their mechanisms on these cell lines. Similar cell death induction and birth inhibition effects of Imatinib and Cetuximab are also observed on most other cell lines in this dataset (Figure S2). In addition, some other compounds have consistent patterns across cell lines. For example, Mebendazole, TCS PIM-11, QNZ and MG-132 demonstrate both birth inhibition and death induction effects on most cell lines. Other compounds like 4-HC, Doxorubicin, Olomoucine II, Valproate, Baicalein, Methylglyoxal and IKK 16 increase cell death rates on most cell lines, but their effects on cell birth rates differ by cell line.

**bdChemo considers uncertainty in chemosensitivity evaluations.** Point estimates of compound potency summary statistics, such as  $GI_{50}$ , on different experiments can be similar even if the growth curves have distinct patterns. In Figure 4a, we plot the lengths of the 95% credible intervals (CI) against the means for  $\log_{10}GI_{50}$  estimated on all experiments from the NCI-DREAM data. There are many cases where means are comparable but lengths of credible intervals differ greatly. For example, on cell line 21NT, compounds Nelfinavir



**Figure 3.** Posterior estimations of Kendall Process means (dose-response curves, left panel), birth rates (middle panel) and death rates (right panel) of cell line BT474 treated with compound Imatinib (a) and cell line HCC1806 treated with compound Cetuximab (b). Curves and shades represent the posterior mean and the 95% credible interval of the corresponding variables, respectively.



**Figure 4.** (a) Length of the credible interval of  $\log_{10}GI_{50}$  v.s. mean of  $\log_{10}GI_{50}$ ; distribution of mean  $\log_{10}GI_{50}$ . (b,c) Dose-response modeling of cell line 21NT treated with compounds Nelfinavir and 4-HC (b), and cell line MDAMB415 treated with compounds Disulfiram and Imatinib (c). Curves and shades in (b,c) represent the posterior means and the 95% credible intervals of the Kendall Process means, respectively. Dashed lines are the  $\log_{10}GI_{50}$ 's calculated from the mean estimations of Kendall Process means.

and 4-HC have  $GI_{50}$  values  $1.33 \times 10^{-5}M$  and  $1.29 \times 10^{-5}M$  respectively, which are close (especially in  $\log_{10}$  scale the difference is negligible), yet Nelfinavir has a 95% CI [ $1.17 \times 10^{-5}, 1.53 \times 10^{-5}$ ]M, much narrower than that of 4-HC, [ $8.87 \times 10^{-6}, 1.74 \times 10^{-5}$ ]M. These differences result from the distinct patterns of their growth curves: compared to the sharp drop of the growth curve under treatment of Nelfinavir (Figure 4b left panel), the growth curve under treatment of 4-HC declines much more slowly as compound concentration increases (Figure 4b right panel), resulting in greater uncertainty about the location of  $GI_{50}$ .

Comparing point estimates of a summary statistic for different compounds can lead to conclusions about their relative potency that lack statistical validity. For instance, on cell line MDAMB415, compounds Disulfiram and Imatinib have  $GI_{50}$  values  $6.55 \times 10^{-6}M$  and  $2.28 \times 10^{-5}M$ , respectively (Figure 4c), based on which, a conclusion may be drawn that MDAMB415 cells are more sensitive to Disulfiram. However, as the  $GI_{50}$  CI of Disulfiram is [ $1.36 \times 10^{-10}, 2.67 \times 10^{-5}$ ]M, covering that of Imatinib, [ $2.00 \times 10^{-5}, 2.62 \times 10^{-5}$ ]M, there is no evidence

supporting the statistical significance of such difference. In practice, summary concentrations are often computed to quickly compare a large number of compounds. However, rigorous statistical hypothesis testing for differences in cell line responses requires taking into account uncertainties in summary statistics used for comparison.

## Discussion

Statistically rigorous analysis of chemosensitivity experiment data is of great importance in cancer therapy development. The major assumptions of the proposed model, bdChemo, based on the birth-death process<sup>13</sup> (BDP), are that the per-cell birth and death rates of a cell line under compound treatment are time-homogenous functions of the compound concentration, and the population-level birth and death rates are products of the cell community size and the per-cell birth and death rates, respectively. Unlike standard analysis methods that rely heavily on specific functional forms of the dose-response curve, bdChemo employs a semi-parametric Bayesian approach in function estimations to avoid restrictive assumptions. Although other nonparametric methods have been proposed for dose-response curve fitting, bdChemo provides biologically motivated estimates of dose-dependence in cell birth and death rates separately, in addition to estimating the combined effects of the compound on cell birth and death processes, delivering richer information that may guide subsequent experimental work. The method takes uncertainty into account when providing chemosensitivity summary statistics, such as  $GI_{50}$ ,  $TGI$  and  $LC_{50}$ , to facilitate sound comparisons of compounds' effects in tumor cell growth inhibition.

We applied bdChemo, as well as two conventional sigmoid-curve fitting approaches (logistic and Gompertz) and R package `grofit`<sup>10</sup>, to NCI-DREAM drug sensitivity prediction challenge<sup>6</sup> data, where dose-response measurements of 28 compounds on 53 breast cancer cell lines were provided. The results show that when the dose-response curve does follow a sigmoid pattern, bdChemo produces estimates similar to conventional methods; but when the dose-response curve deviates from a sigmoidal shape, bdChemo and nonparametric spline smoothing can better capture growth inhibition dynamics. We observe different patterns of cell birth inhibition and death induction effects for different cell line/compound combinations; while some compound-cell line combinations have dose-response curves that look similar, the underlying mechanism of action may be different. Separate estimates of cell birth and death rate estimations provided by bdChemo, hence, might be utilized in hypothesis generation and experimental design. In addition, even when the point estimates of  $GI_{50}$  are similar, credible intervals sometimes differ substantially.

While bdChemo can fit curves flexibly and provides mechanistic inferences about the dose-dependent response of cell birth and death rates, the approach is subject to limitations. First, the Kendall process framework assumes that cells undergo birth and death independently, with the same per-cell rates; when there are  $k$  cells, the population-level birth and death rates are  $k\lambda(z)$  and  $k\mu(z)$  respectively. The model does not accommodate population-level effects, which could result in more complicated nonlinear rates  $\lambda(k,z)$  and  $\mu(k,z)$ . More sophisticated models for population effects of cell response may be warranted when biologically motivated. Second, we analyzed each experiment (compound-cell line combination) independently. Jointly modeling inhibition responses by compound and cell line could exploit information sharing across experiments, resulting in greater statistical precision in estimates, especially when the number of doses is small.

Finally, we point out potential issues that may arise when applying bdChemo in empirical data analysis. We observed some large credible intervals of  $\lambda$  and  $\mu$  in our analysis, which is mainly caused by the small number of experimental replicates (three at each dose). Since the mean value of the cell numbers only contains information regarding the combined effects of the birth and death processes, the separate identifiability of the birth and death rates mainly comes from the variance of cell numbers at different doses. Therefore, a larger number of replicates is desirable for increased accuracy in estimates of  $\lambda$  and  $\mu$ . Additionally, we observed a few outliers in which  $\lambda$  and  $\mu$  are estimated to be much larger than elsewhere (Tables S2 and S3). These estimates are driven by large cell count variations in these experiments. For a comprehensive evaluation of our method's performance on datasets of varying qualities, we analyzed all NCI-DREAM experiments here. However, in practice, some quality control procedures in data preprocessing, as typically conducted in conventional chemosensitivity analysis<sup>18</sup>, might also be necessary.

## Methods

**Conventional approaches.** Conventionally a cell line's response to a compound is modeled with a sigmoid curve. The most commonly used include Gompertz curve<sup>5,6</sup>

$$g(z) = \varphi_0 + \varphi_1 \exp(-\exp(\varphi_2 + \varphi_3 z)), \quad (1)$$

and logistic curve<sup>7,8</sup>

$$g(z) = \psi_0 + \frac{\psi_1}{1 + \exp\left(\frac{z - \psi_2}{\psi_3}\right)}, \quad (2)$$

where  $g(z)$  is the final cell count at  $\log_{10}$  concentration,  $z$  (usually with unit  $\log_{10} M$ ).

**bdChemo. Model.** We use birth-death process (BDP) to model cell growth in the experiment. In a general BDP<sup>19</sup>, where  $N(t)$  stands for the number of particles at time  $t$ , given  $N(t) = k (k \geq 1)$ , the birth rate

$$\lambda_k = \lim_{dt \rightarrow 0} \frac{\Pr(N(t + dt) = k + 1 | N(t) = k)}{dt} \quad (3)$$



and the death rate

$$\mu_k = \lim_{dt \rightarrow 0} \frac{\Pr(N(t+dt) = k-1 | N(t) = k)}{dt} \quad (4)$$

are time-homogeneous but dependent on the number of particles  $k$ . A simple linear BDP known as Kendall Process assumes  $\lambda_k = \lambda k$  and  $\mu_k = \mu k$  (we refer  $\lambda$  and  $\mu$  as per cell birth and death rates, respectively)<sup>13</sup>. For Kendall Process, the transition probability is<sup>20</sup>

$$\begin{aligned} P_{ab}(t) &= \Pr(N(t) = b | N(0) = a) \\ &= \sum_{j=0}^{\min(a,b)} \binom{a}{j} \binom{a+b-j-1}{a-1} \alpha^{a-j} \beta^{b-j} (1-\alpha-\beta)^j, \end{aligned} \quad (5)$$

where

$$\alpha(t) = \frac{\mu(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu} \quad (6)$$

and

$$\beta(t) = \frac{\lambda(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu} \quad (7)$$

For a fixed  $t$ ,  $N(t)$  is a random variable whose distribution is fully specified by  $P_{ab}(t)$ . The mean and variance of  $N(t)$  given  $N(0) = n_0$  are<sup>19,21</sup>

$$m_t = n_0 \exp((\lambda - \mu)t) \quad (8)$$

and

$$v_t = n_0 \frac{\lambda + \mu}{\lambda - \mu} \exp((\lambda - \mu)t) [\exp((\lambda - \mu)t) - 1], \quad (9)$$

respectively.

As indicated above, to calculate  $P_{ab}(t)$  involves computing a large number of combinatorial numbers, making it computationally infeasible in real applications<sup>22</sup>. Therefore, we use normal distribution with matched mean ( $m_t$ ) and variance ( $v_t$ ) as an approximation to reduce the computational cost. This approximation is accurate when initial cell counts are large, as they generally are in chemosensitivity experiments (see the Supplementary Note for a technical justification). Since in one chemo-sensitivity study, treatment duration  $t$  is usually fixed and same for all experiments (compounds), we omit the notation  $t$  for simplicity and interpret the new  $\lambda$  and  $\mu$  as per cell birth and death rates for the entire experiment duration.

For a given compound working on a given cell line, we assume  $\lambda$  and  $\mu$  are functions of  $\log_{10}$  concentration,  $z$ , of the compound, so the mean and variance of cell counts are

$$m(n_0, z) = n_0 \exp(\lambda(z) - \mu(z)), \quad (10)$$

and

$$v(n_0, z) = n_0 \frac{\lambda(z) + \mu(z)}{\lambda(z) - \mu(z)} \exp(\lambda(z) - \mu(z)) [\exp(\lambda(z) - \mu(z)) - 1], \quad (11)$$

respectively.

Note, although the mean as a function of  $\lambda$  and  $\mu$  is only affected by their difference, the variance has a term  $\lambda + \mu$ . Therefore, mean and variance of cell counts together provides information to identify  $\lambda$  and  $\mu$  separately. An interesting example is illustrated by the difference between quiescence ( $\lambda \approx 0$  and  $\mu \approx 0$ ) and matched birth and death rates ( $\lambda \approx \mu$ ). In both cases, we would expect the mean cell count to be unchanged,  $m(n_0, z) \approx n_0$ . However, if both  $\lambda$  and  $\mu$  are close to 0, the variance  $v(n_0, z)$  would also be very small; if the rates are large and nearly equal,  $\lambda \approx \mu \gg 0$ , the variance is expected to be much larger.

Our model also takes experimental errors into account by assuming that the measured cell counts  $X_z = N_z + \epsilon_z$  and  $\epsilon_z \sim N(\theta, \sigma^2)$ , where  $\theta$  and  $\sigma^2$  are the mean and variance of background noises, respectively.

Therefore, the likelihood of observing data  $D = \{(n_{01}, z_1, x_1), (n_{02}, z_2, x_2), \dots, (n_{0n}, z_n, x_n), e_1, \dots, e_q\}$  ( $n_{0i}$ 's are the initial cell population sizes,  $(z_i, x_i)$ 's are compound concentrations and corresponding cell count measurements at follow-up, and  $e_i$ 's are independent background noise measurements) is

$$L(D|\lambda, \mu, \sigma^2) = \prod_{i=1}^n f_N(x_i | m(n_{0i}, z_i) + \theta, v(n_{0i}, z_i) + \sigma^2) \prod_{i=1}^q f_N(e_i | \theta, \sigma^2), \quad (12)$$

where  $f_N$  is the density function of Gaussian distribution.

Note that we do not require the initial cell population sizes  $n_{0i}$  to be the same across different compound concentrations, but we treat them as known quantities here. In reality, initial cell population sizes may be uncertain in some experiments. However, with only one number provided in most experiment data, it is difficult to assign a distribution to the initial cell count. Moreover, uncertainty in the initial count mainly comes from variation in cell density and the amount of cell solution injected into each well, both of which, under appropriate experimental operation, could be controlled at a relatively low level. Although we do not model its randomness directly, the term  $\sigma^2$  models the variance from background noise can be interpreted to reflect variation in initial cell counts to some extent. For example, if two datasets have all other conditions similar, whereas one has a much larger estimate of  $\sigma^2$ , we might suspect that the large  $\sigma^2$  is induced by a poorly controlled initial cell seeding. If future experiments provide more data to quantify its fluctuations and suggest the necessity of treating it as a random variable, we may add another layer of randomness to  $n_0$  under this Bayesian hierarchical model framework.

To avoid unnecessary assumptions on  $\lambda(z)$  and  $\mu(z)$  and allow more flexible estimations, we employ a semi-parametric Bayesian approach by assigning Gaussian process<sup>14</sup> priors to  $\phi_\lambda(z) = \log(\lambda(z))$  and  $\phi_\mu(z) = \log(\mu(z))$ , i.e.

$$\phi_\lambda(z) \sim GP(\alpha_\lambda, K_\lambda), \quad (13)$$

$$\phi_\mu(z) \sim GP(\alpha_\mu, K_\mu), \quad (14)$$

and

$$K_\lambda(z, z') = \tau_\lambda^2 \exp\left(-\frac{(z - z')^2}{l_\lambda^2}\right), \quad (15)$$

$$K_\mu(z, z') = \tau_\mu^2 \exp\left(-\frac{(z - z')^2}{l_\mu^2}\right). \quad (16)$$

Note, unlike conventional approaches, under this Gaussian process framework, we impose no shape constraints including monotonicity on the curves.

Hyperparameters  $\alpha_\lambda$ ,  $\alpha_\mu$ ,  $\tau_\lambda^2$ ,  $\tau_\mu^2$ ,  $l_\lambda^2$ ,  $l_\mu^2$ ,  $\theta$  and  $\sigma^2$  are assigned priors

$$\alpha_\lambda \sim N(0, sa_\lambda^2), \quad (17)$$

$$\alpha_\mu \sim N(0, sa_\mu^2), \quad (18)$$

$$\tau_\lambda^2 \sim \text{Inverse} - \text{Gamma}(a_{1\lambda}, b_{1\lambda}), \quad (19)$$

$$\tau_\mu^2 \sim \text{Inverse} - \text{Gamma}(a_{1\mu}, b_{1\mu}), \quad (20)$$

$$l_\lambda^2 \sim \text{Gamma}(a_{2\lambda}, b_{2\lambda}), \quad (21)$$

$$l_\mu^2 \sim \text{Gamma}(a_{2\mu}, b_{2\mu}), \quad (22)$$

$$\theta \sim \text{Unif}(-\infty, \infty), \text{ i. e. } p(\theta) \propto 1 \quad (23)$$

and

$$\sigma^2 \sim \log - \text{Unif}(0, 1), \text{ i. e. } p(\sigma^2) \propto \frac{1}{\sigma^2}. \quad (24)$$

**Fitting algorithm.** We utilize Gibbs sampler embedded with Metropolis updating<sup>15</sup> to draw posterior samples for parameters  $\phi_\lambda$ ,  $\phi_\mu$ ,  $\alpha_\lambda$ ,  $\alpha_\mu$ ,  $\tau_\lambda^2$ ,  $\tau_\mu^2$ ,  $l_\lambda^2$ ,  $l_\mu^2$ ,  $\theta$  and  $\sigma^2$  (see Supplementary Note for details).

**Data availability.** The dataset analyzed during the current study is available at <http://www.nature.com/nbt/journal/v32/n12/full/nbt.2877.html>. The code used to perform the analysis is available as an R package “bdChemo” at <https://github.com/YiyiLiu1/bdChemo>.

## References

1. Rubin, E. H. & Gilliland, D. G. Drug development and clinical trials—the path to an approved cancer drug. *Nat Rev Clin Oncol* **9**, 215–222 (2012).
2. Mayr, L. M. & Bojanic, D. Novel trends in high-throughput screening. *Curr Opin Pharmacol* **9**, 580–588 (2009).

3. Riss, T. L. *et al.* In *Assay Guidance Manual*. (eds G.S. Sittampalam *et al.*) (Bethesda (MD); 2004).
4. Paul, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* **9**, 203–214 (2010).
5. Winsor, C. P. The Gompertz curve as a growth curve. *Proceedings of the National Academy of Sciences of the United States of America* **18**, 1–8 (1932).
6. Costello, J. C. *et al.* A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol* **32**, 1202–1212 (2014).
7. Pearl, R. & Reed, L. J. On the Rate of Growth of the Population of the United States since 1790 and Its Mathematical Representation. *Proc Natl Acad Sci USA* **6**, 275–288 (1920).
8. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
9. Monks, A. *et al.* Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines. *J Natl Cancer Inst* **83**, 757–766 (1991).
10. Kahm, M., Hasenbrink, G., Lichtenberg-Frate, H., Ludwig, J. & Kschischo, M. grofit: Fitting Biological Growth Curves with R. *J Stat Softw* **33**, 1–21 (2010).
11. Schenone, M., Dancik, V., Wagner, B. K. & Clemons, P. A. Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol* **9**, 232–240 (2013).
12. Kwong, L. N. *et al.* Oncogenic NRAS signaling differentially regulates survival and proliferation in melanoma. *Nat Med* **18**, 1503–1510 (2012).
13. Kendall, D. G. On the Generalized Birth-and-Death Process. *Annals of Mathematical Statistics* **19**, 1–15 (1948).
14. Rasmussen, C. E. & Williams, C. K. I. *Gaussian processes for machine learning*. (MIT Press, Cambridge, Mass.; 2006).
15. Gelman, A. *Bayesian data analysis*, Edn. 3rd. (Chapman & Hall/CRC, Boca Raton, Fla.; 2014).
16. Goldman, J. M. & Melo, J. V. Chronic myeloid leukemia—advances in biology and new approaches to treatment. *N Engl J Med* **349**, 1451–1464 (2003).
17. Karapetis, C. S. *et al.* K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* **359**, 1757–1765 (2008).
18. Heiser, L. M. *et al.* Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc Natl Acad Sci USA* **109**, 2724–2729 (2012).
19. Feller, W. *An introduction to probability theory and its applications*, Edn. 3d. (Wiley, New York, 1971).
20. Crawford, F. W. & Suchard, M. A. Transition probabilities for general birth-death processes with applications in ecology, genetics, and evolution. *Journal of Mathematical Biology* **65**, 553–580 (2012).
21. Renshaw, E. *Stochastic population processes: analysis, approximations, simulations*. (Oxford University Press, Oxford; New York; 2011).
22. Crawford, F. W., Minin, V. N. & Suchard, M. A. Estimation for General Birth-Death Processes. *Journal of the American Statistical Association* **109**, 730–747 (2014).

## Acknowledgements

We are grateful to Dr. Hongyu Zhao, Dr. David Stern, Dr. Christos Hatzis, Li Zeng and Michael Klein for helpful discussions and suggestions for the manuscript. F.W.C. was supported by NIH grant NICHD 1DP2HD091799-01.

## Author Contributions

Y.L. and F.W.C. conceived the research. Y.L. implemented the method and performed the data analysis. Y.L. and F.W.C. wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-21017-5>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018