Inferring Drug-induced Plasticity via Drug Screen Data

Chenyu Wu, Einar Bjarki Gunnarsson, Jasmine Foo, Kevin Leder

September 30, 2024

Abstract

Resistance to therapy remains a significant challenge in cancer treatment, often due to the presence of a stem-like cell population that drives tumor recurrence post-treatment. Moreover, many anticancer drugs inadvertently induce plasticity, reverting differentiated cancer cells to drug-resistant stem-like states. Addressing this phenomenon is crucial for advancing cancer therapeutics. In this study, we introduce a robust statistical framework based on multi-type branching processes that can dissect tumor dynamics and drug effects from high throughput drug screening data. Through comprehensive *in silico* experiments, we show the efficacy of our framework in estimating parameters governing population dynamics and drug responses in a heterogeneous tumor population where cell state transitions are influenced by the drug. Finally, using recent *in vitro* data involving AGS-SORE6+/− cells treated with ciclopirox olamine, we show how our framework can be used to confirm the presence of drug-induced cell plasticity.

1 Introduction

Cellular plasticity, the potential of immature cells to differentiate into various mature terminal cells, is a fundamental cell feature. This process is reversible and can be induced by external signals [34]. In the context of cancer, cellular plasticity is often observed during treatment, a phenomenon known as drug-induced plasticity. Specifically, it refers to the process where a more mature cancer cell phenotype, a cancer non-stem-like cell (CNSC), reverts to a more immature cancer stem-like phenotype (CSC) in response to therapy. This phenomenon has been observed under both pharmaceutical treatments and other treatment modalities such as radiotherapy [17, 12, 31].

Drug-induced plasticity is thought to play a significant role in the evolution of drug resistance [31, 2]. In particular, CSCs can confound treatment efforts due to their inherent resistance to therapy and their ability for unlimited cell division [3]. As such, understanding drug-induced transitions from CNSCs to CSCs can help shed light on tumor heterogeneity and aid in the development of improved treatment plans [18]. In addition, since Yamanaka et al. demonstrated the generation of induced pluripotent stem cells from human dermal fibroblasts [34], artificially enhancing cellular plasticity has gained popularity in cancer research. By enhancing cellular plasticity, researchers can reprogram cancer cells and facilitate the investigation of tumorigenesis [13, 30]. Small-molecule drugs represent one promising avenue for artificially reprogramming differentiated cells [15].

In a recent study [25], ciclopirox olamine (CPX-O), originally an antifungal agent, was identified for its ability to induce the production of gastric CSCs from gastric CNSCs. The study utilized the SORE-GFP reporter system coupled with fluorescence-activated cell sorting (FACS) to distinguish between gastric CSCs and CNSCs. A high-throughput screening (HTS) technique was employed to assess the effects of various potential medications on a homogeneous population of gastric CNSC to explore drug-induced plasticity. One key assumption of this study is the effective separation of

gastric CSCs and CNSCs using the FACS technique and the SORE-GFP reporter system, which may limit the scope of the findings. Additionally, the study noted CPX's significant cytotoxic impact on gastric CNSCs. This complicates assessment of the level of drug-induced plasticity, since increased abundance of CSCs relative to CNSCs can both be explained by drug-induced plasticity and a selective advantage over CNSCs. Consequently, the empirical inference of drug-induced plasticity presents significant challenges. As an alternative approach, we propose using a mathematical model to understand how the drug affects the proliferation of both CNSCs and CSCs, and to tease apart the drug's effect on cell proliferation vs. transitions from CNSCs to CSCs.

Mathematical modeling has increasingly played a crucial role in understanding and treating cancer [38, 20], offering advantages in modeling complex biological dynamics through manageable mathematical models. For instance, recent studies have applied the multi-type branching process model to depict diverse tumor populations undergoing phenotypic switching [14, 8, 7]. This approach abstracts the cell division process into 'branching events', allowing each cell to generate descendants across any subtype described in the model. As a result, the model can naturally describe processes such as self-proliferation, differentiation, and de-differentiation of CSCs and CNSCs.

On the other hand, several recent studies [16, 35] have proposed frameworks for deconvoluting subpopulation structure directly from HTS bulk cell count data, avoiding the need for a reporter system coupled with FACS to separate distinct subpopulations. Instead, these studies distinguish subpopulations based on their distinct responses to a given drug. For instance, these works have successfully identified Imatinib-sensitive and -resistant Ba/F3 cells from a population mixture by examining heterogeneous drug response curves at various concentration levels. However, these frameworks assume that each subpopulation can only generate its own subtype, thereby limiting their ability to capture the differentiation and de-differentiation dynamics of CSCs and CNSCs.

In this study, we have developed a novel statistical framework that integrates the multi-type branching process with a drug response model to simulate heterogeneous cell growth dynamics under drug influence. This newly proposed framework serves as a mathematical tool for investigating drug-induced plasticity directly from HTS bulk cell count data. The paper is organized as follows. In Section 2, we introduce our newly proposed framework. Based on this framework, we conduct multiple *in silico* experiments to validate our model predictions in Section 3. Subsequently, in Section 4, we validate our framework using *in vitro* data obtained from [25] and [16]. Finally, in Section 5, we conclude by discussing the advantages and limitations of our framework. A detailed derivation of the methods and the experimental settings are provided in the Appendix.

2   Methodology

In this section, we start by introducing a general model to describe heterogeneous tumor growth dynamics, capable of accommodating an arbitrary number of distinct phenotypes. Then, we develop a drug-effect model using the classical Hill equation, which is widely employed to represent dose-response relationships. By integrating these two models, we propose a novel statistical framework for inferring underlying tumor dynamics from HTS data. Finally, we tailor this framework to a specific case involving two subpopulations within the tumor, CSCs and CNSCs, which will be the focus of our analysis in the remainder of the paper. In what follows, we denote vectors and matrices in boldface letters and all the vectors are row vectors.

2.1   Asymmetrical birth multi-type branching process model

Assume there are K subpopulations of phenotypes within a tumor, each exhibiting its own growth dynamics and drug response. These subpopulations are indexed by $\mathcal{K} = \{1, \cdots, K\}$. Increasing

2

evidence indicates that asymmetric cell division is a key mechanism for both differentiation [11, 22] and dedifferentiation [32]. Therefore, we assume that each phenotype can transition to one or more other phenotypes via asymmetric cell division. Our framework can also readily accommodate phenotypic transitions that occur independently of cell divisions, as discussed for example in [8].

We employ an asymmetric birth multi-type branching process model in continuous time [1]. In the model, a type-i cell divides into two type-i cells at rate α_i, experiences natural death at rate β_i, and asymmetrically divides into one type-i cell and one type-j cell at rate ν_{ij} (see Figure 1). Specifically, within a given infinitesimal time interval $\Delta t > 0$, a type-i cell has a probability $\alpha_i \Delta t$ of dividing into two type-i cells, $\beta_i \Delta t$ of dying, and $\nu_{ij} \Delta t$ of asymmetrically dividing into one type-i and one type-j cell. The net growth rate of type-i cells, κ_i, is defined as $\kappa_i := \alpha_i - \beta_i$. The dynamics of the model are encoded in the *infinitesimal generator matrix* $\mathbf{A}$, which is the $K \times K$ matrix

$$\mathbf{A} := \begin{bmatrix} \kappa_1 & \nu_{12} & \cdots & \nu_{1K} \\ \nu_{21} & \kappa_2 & \cdots & \nu_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \nu_{K1} & \nu_{K2} & \cdots & \kappa_K \end{bmatrix}. \tag{1}$$

In this matrix, the (i, j)-th element is the net rate at which a type-i cell produces a type-j cell. For all K phenotypes, we assume a strictly positive birth rate ($\alpha_i > 0, i \in \mathcal{K}$), a non-negative death rate ($\beta_i \geq 0, i \in \mathcal{K}$), and a non-negative asymmetric birth rate ($\nu_{i,j} \geq 0, i \neq j$). We furthermore assume that $\kappa_i > 0$ for all $i \in \mathcal{K}$, i.e. the net birth rate is positive, though transition and death may not occur for all phenotypes.

For simplicity, we assume that the multi-type branching process is *irreducible*, meaning that each subpopulation can eventually produce descendants in any other subpopulation, possibly through intermediate types. Mathematically, this means that the infinitesimal generator matrix $\mathbf{A}$ cannot be transformed into a block upper triangular matrix via simultaneous row or column permutation.

(a) Symmetric division (rate: α_1)   (b) Death (rate: β_1)   (c) Asymmetric division (rate: ν_{12})
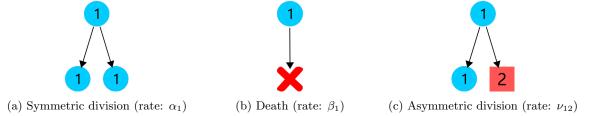
Figure 1: Illustration of three independent events in the asymmetrical birth multi-type branching process model. Different shapes represent different phenotypes.

We denote the cell counts of each phenotype as a random vector: $\mathbf{B}(t) = [B_1(t), B_2(t), \cdots, B_K(t)]$, where $B_i(t)$ is the number of type-i cells at the time t. For the special case where the process is started by a single type-i cell, the cell count vector at time t is denoted by $\mathbf{B}^{(i)}(t) = [B_1^{(i)}(t), B_2^{(i)}(t), \cdots, B_K^{(i)}(t)]$, where $\mathbf{B}^{(i)}(0) = \mathbf{e}_i$ is the i-th identity vector. The *mean vector* and *covariance matrix* for $\mathbf{B}^{(i)}(t)$ are denoted by

$$\mathbf{m}^{(i)}(t) := \mathrm{E}\left[\mathbf{B}^{(i)}(t)\right],$$

$$\mathbf{\Xi}^{(i)}(t) := \mathrm{E}\left[\left(\mathbf{B}^{(i)}(t) - \mathbf{m}^{(i)}(t)\right)^{\mathsf{T}} \left(\mathbf{B}^{(i)}(t) - \mathbf{m}^{(i)}(t)\right)\right],$$

3

where $t \geq 0$ and $\left(\mathbf{B}^{(i)}(t) - \mathbf{m}^{(i)}(t)\right)^{\mathsf{T}}$ represents the transpose of the vector $\left(\mathbf{B}^{(i)}(t) - \mathbf{m}^{(i)}(t)\right)$. To explicitly compute the *mean vector*, we define the *mean matrix* using the matrix exponential:

$$\mathbf{M}(t) = \exp\left(t\mathbf{A}\right) = \sum_{j=0}^{\infty} \left(\frac{t^j}{j!}\right) \mathbf{A}^j.$$

The *mean vector* $\mathbf{m}^{(i)}(t)$ can be computed as the i-th row of the *mean matrix*,

$$\mathbf{m}^{(i)}(t) = \mathbf{e}_i \mathbf{M}(t),$$

and the covariance matrix $\mathbf{\Xi}^{(i)}(t)$ can be computed as shown in Proposition 1 below.

2.2 Drug-effect Model

Before modeling the drug effect, we introduce a fixed drug dose d and denote the cell count at time t and concentration level d as $\mathbf{B}(t, d)$. The *mean matrix* $\mathbf{M}(t, d)$ and the covariance matrix $\mathbf{\Xi}^{(i)}(t, d)$ follow. Note that we use concentration level and dose interchangeably.

2.2.1 Base model

To model the drug effect, we define the well-known Hill equation with parameters (b, E, m) as

$$H(d; b, E, m) = b + \frac{1 - b}{1 + (d/E)^m}.$$

The Hill equation is a classic sigmoidal function used to describe a dose-response [27]. In the equation, the parameter b represents the maximum drug effect, since $H(0; b, E, m) = 1$ and $\lim_{d\to\infty} H(d; b, E, m) = b$. For $b \in (0, 1)$, the Hill equation is strictly decreasing, while it is strictly increasing for $b > 1$ (Figure 2). The parameter E indicates the concentration at which 50 percent of the maximum effect is attained, known as the inflection point. The last parameter m is the Hill coefficient, which controls the steepness of the Hill equation around the inflection point. As a simplification, we fix the Hill parameter $m = 1$ for all drug effects when conducting our *in silico* experiments in Section 3. We then reintroduce this parameter when modeling *in vitro* data in Section 4.

This study focuses on the time-homogeneous effects of a single drug. We distinguish between two types of drug responses, drug toxicity response and drug-induced plasticity response, thereby capturing multiple effects from a single drug.

- *Drug toxicity response (cytotoxic effect)*: The parameters related to this response are denoted as $(b_{i,\beta}, E_{i,\beta})$ for type i. We assume that $b_{i,\beta} \in (0, 1)$ and that the death rate of type-i under dose d is given by (for the case $m = 1$):

$$\beta_i(d) = \beta_i - \log(H(d; b_{i,\beta}, E_{i,\beta})) = \beta_i - \log\left(b_{i,\beta} + \frac{1 - b_{i,\beta}}{1 + (d/E_{i,\beta})}\right). \tag{2}$$

  The net growth rate of a type i cell, κ_i, is therefore negatively affected by increasing drug concentration, i.e. $\kappa_i(d) = (\alpha_i - \beta_i) + \log(H(d; b_{i,\beta}, E_{i,\beta}))$. In this scenario, the drug is assumed to act through a cytotoxic mechanism, meaning that higher doses result in increased rates of cell death. We note that our framework can also easily incorporate cytostatic effects, where higher doses lead to a lower cell division rate, using a similar Hill function.
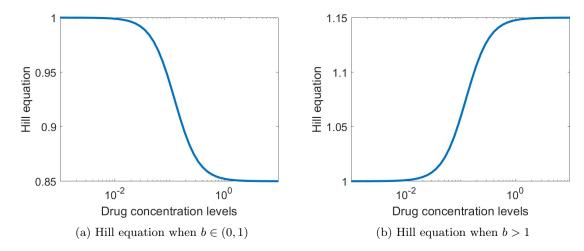
(a) Hill equation when $b \in (0, 1)$ (b) Hill equation when $b > 1$

Figure 2: Hill equation illustrations

- *Drug-induced plasticity response*: The parameters related to this response are denoted as $(b_{i,\nu}, E_{i,\nu})$ for type-i. For simplicity, we assume here that the drug effect on type-i cell transitions is independent of the target phenotype j. Drug-induced phenotypic transition is a widely observed effect of cancer treatment [26, 6, 33]. Particularly, drug-sensitive phenotypes may be induced to develop resistance when exposed to the drug. When modeling the drug effect on the asymmetric birth rate, we assume $b_{i,\nu} \geq 1$ to account for elevated transitions between subpopulations caused by the therapeutic environment. We furthermore assume that the asymmetric birth rate of type i to any other type j under dose d is given by (for the case $m = 1$):

$$\nu_{ij}(d) = \nu_{ij} + \log(H(d; b_{i,\nu}, E_{i,\nu})) = \nu_{ij} + \log\left(b_{i,\nu} + \frac{1 - b_{i,\nu}}{1 + (d/E_{i,\nu})}\right). \tag{3}$$

The assumption $b_{i,\nu} \geq 1$ incorporates the situation where the drug does not impact phenotypic transitions, when $b_{i,\nu} = 1$, $\nu_{ij}(d) = \nu_{ij}$ for all $d \geq 0$. It is worth mentioning that our framework can also model drug inhibition of the asymmetric birth rate by letting $b_{i,\nu} \in (0, 1)$.

2.2.2 Logistic time-delayed drug response

Our framework can be extended to accommodate a more complex drug-effect model, which may involve time-inhomogeneous drug effects. In particular, for the *in vitro* dataset from [25] investigated in Section 4, there appear to be delayed drug effects, which have been observed in other studies as well [10, 29, 36]. To address this behavior, we allow for the possibility of a time-inhomogeneous drug effect, which is modeled using a two-parameter logistic function. In this case, we assume that the maximum drug effect parameter b depends on time t through

$$|1 - b(t)| = \frac{1}{1 + \exp(k(t_0 - t))}|1 - b|,$$

where k, t_0 are logistic function parameters. In this way, we can model the drug's maximum potential $b(t)$ gradually reaching its ultimate value as time increases. Absolute values are taken to allow both the cytotoxic effect $b(t) < 1$ and the drug-induced plasticity effect $b(t) > 1$ to be time-inhomogeneous.

2.3 Long-run behavior

As the HTS experiments being considered are run for a longer time duration, it becomes important to understand the long-run behavior of the process $\mathbf{B}(t, d)$. According to the Perron-Frobenius Theorem, the irreducible infinitesimal generator $\mathbf{A}(d)$ has a strictly positive largest eigenvalue $\lambda_1(d)$, which corresponds to the largest eigenvalue $e^{\lambda_1(d)t}$ of the mean matrix $\mathbf{M}(t, d)$. Additionally, the corresponding left eigenvector $\boldsymbol{\pi}(d)$ is strictly positive. A well-known limiting result reviewed in [1] states that there exists, almost surely, a non-negative numerical random variable W such that

$$\lim_{t \to \infty} \mathbf{B}(t, d) e^{-\lambda_1(d)t} = W \boldsymbol{\pi}(d).$$

This result characterizes two aspects of the long-run behavior of the process $\mathbf{B}(t, d)$. First, the stochastic process $\mathbf{B}(t, d)$ will grow with a deterministic exponential *long-run growth rate* $\lambda_1(d)$. Second, when the eigenvector $\boldsymbol{\pi}(d)$ is normalized so that $\sum_{i=1}^{K} \pi_i(d) = 1$, then $\pi_i(d)$ describes the long-run proportion of type-i cells in the population. Therefore, we refer to the normalized version of $\boldsymbol{\pi}(d)$ as the *stable proportion* between phenotypes.

2.4 Statistical Model

Now, we construct a statistical framework to infer heterogeneous tumor growth dynamics and drug responses using HTS live-cell imaging bulk data. Here, 'bulk data' refers to aggregated total cell counts across all subpopulations within the heterogeneous tumor. Our framework is designed to accommodate data collected via live-cell imaging techniques, allowing efficient capture of bulk cell counts from a single sample across multiple time points [24].

Assume the bulk dataset $\mathbf{X}$ is collected across a set of drug concentration levels $\mathcal{D} = \{d_1, \cdots, d_{N_D}\}$ and a set of time points $\mathcal{T} = \{t_1, \cdots, t_{N_T}\}$, where $0 < t_1 < \cdots < t_{N_T}$. For each drug dose $d \in \mathcal{D}$, N_R samples are cultivated, and live-cell imaging technique is used to collect bulk cell count data at time points $t_1, \ldots, t_{N_T}$. Let $\mathbf{x}_{\mathcal{T}, d, r} = (x_{d, r, 1}, \ldots, x_{d, r, N_T})$ denote the data collected for the r-th replicate under drug concentration d, where $x_{d, r, k}$ is the bulk cell count at time point t_k. Taking into account experimental measurement error, we propose the statistical model

$$\mathbf{x}_{\mathcal{T}, d, r} = \mathbf{Y}_{d, r}(\mathcal{T}) + \mathbf{Z}_{d, r}, \tag{4}$$

where $\mathbf{Y}_{d, r}(\mathcal{T})$ denotes bulk cell counts at the time points $\mathcal{T}$ under the stochastic model outlined in Sections 2.1 and 2.2, and $\mathbf{Z}_{d, r} \sim N(0, c^2 \mathbf{I})$ are independent multivariate normally distributed noise terms. Assuming that n_i is the starting number of cells of type i in each experiment, we can write

$$\mathbf{Y}_{d, r}(\mathcal{T}) := \sum_{i=1}^{K} \sum_{j=1}^{n_i} \mathbf{y}_j^{(i)}(\mathcal{T}, d, r), \tag{5}$$

where $\mathbf{y}_j^{(i)}(\mathcal{T}, d, r)$ denotes the size of the clone started by the j-th initial type i cell at the time points $\mathcal{T}$ under the stochastic model. Let n denote the initial total cell count and $\mathbf{p} = [p_1, p_2, \cdots, p_K]$ denote the initial subpopulation distribution, with $n_i = np_i$ for all i. In this study, we assume that p_i is independent of n for all i, implying that $n_i \to \infty$ as $n \to \infty$.

2.4.1 Central limit theorem

For simplicity, we temporarily omit the explicit notation for drug concentration levels and experimental replicates and write $\mathbf{Y}(\mathcal{T})$ and $\mathbf{y}_j^{(i)}(\mathcal{T})$. When bulk cell counts are observed as opposed

to individual subpopulation counts, the stochastic model outlined in Section 2.1 and 2.2 is no longer a Markov process. Therefore, the exact probability distribution for $\mathbf{Y}(\mathcal{T})$ becomes overly complex [35]. Given that drug screening experiments are usually started by a relatively large number of cells (in [4] for example, 2500 cells were deployed in each well), it is natural to focus on the asymptotic behavior of $\mathbf{Y}(\mathcal{T})$ as the initial cell number increases ($n \to \infty$). For that, we need to understand the asymptotic behavior of $\sum_{j=1}^{n_i} \mathbf{y}_j^{(i)}(\mathcal{T})$. We denote the expectation

$$\mathrm{E}\left[\mathbf{y}_j^{(i)}(\mathcal{T})\right] := \boldsymbol{\mu}^{(i)}(\mathcal{T}) = [\langle \mathbf{m}^{(i)}(t_1), \mathbf{1}\rangle, \cdots, \langle \mathbf{m}^{(i)}(t_{N_T}), \mathbf{1}\rangle],$$

where $\mathbf{1}$ is a size K all ones vector, $\langle \cdot, \cdot \rangle$ is the inner product, and $\mathbf{m}^{(i)}(t)$ denotes the mean vector described in Section 2.1. We furthermore define a centered and normalized process for each $i \in \mathcal{K}$ by

$$\mathbf{W}_{n_i}^{(i)}(\mathcal{T}) := \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \left(\mathbf{y}_j^{(i)}(\mathcal{T}) - \boldsymbol{\mu}^{(i)}(\mathcal{T})\right).$$

A direct application of the multivariate central limit theorem gives the following result, where '$\Rightarrow$' denotes converge in distribution.

Proposition 1. *As $n_i \to \infty$:*
$$\mathbf{W}_{n_i}^{(i)}(\mathcal{T}) \Rightarrow N(\mathbf{0}, \mathbf{V}^{(i)}(\mathcal{T}))$$

where
$$\mathbf{V}_{a,b}^{(i)}(\mathcal{T}) = \mathbf{1}\boldsymbol{\Xi}^{(i)}(t_a)\mathbf{M}(t_b - t_a)\mathbf{1}^{\intercal} \quad for\ 1 \le a \le b \le N_T.$$

Here, $\mathbf{M}(t)$ is the mean matrix, and $\boldsymbol{\Xi}^{(i)}(t_a)$ is a covariance matrix (Section 2.1) given by

$$\boldsymbol{\Xi}^{(i)}(t) = \mathrm{diag}(\mathbf{m}^{(i)}(t)) - \mathbf{m}^{(i)}(t)^{\intercal}\mathbf{m}^{(i)}(t) + \int_0^t (\mathbf{M}(t - \tau))^{\intercal}\, \mathbf{C}^{(i)}(\tau)\, (\mathbf{M}(t - \tau))\, d\tau,$$

where
$$\mathbf{C}_{jk}^{(i)}(\tau) = \begin{cases} \nu_{kj}\mathbf{M}_{ik}(\tau) + \nu_{jk}\mathbf{M}_{ij}(\tau) & j \ne k \\ 2\alpha_j\mathbf{M}_{ij}(\tau) & j = k. \end{cases}$$

This result can be derived from a more general result, which is stated as Proposition 2 in Appendix 6.4. According to Proposition 1, we can approximate $\sum_{j=1}^{n_i} \mathbf{y}_j^{(i)}(\mathcal{T})$ as

$$\sum_{j=1}^{n_i} \mathbf{y}_j^{(i)}(\mathcal{T}) \approx n_i\boldsymbol{\mu}^{(i)}(\mathcal{T}) + N(0, n_i\mathbf{V}^{(i)}(\mathcal{T}))$$

for sufficiently large n_i. We can approximate $\mathbf{Y}(\mathcal{T})$ accordingly. It is important to note that this approximation holds true for each fixed drug concentration level d.

2.4.2 Maximum likelihood estimation (MLE) framework

Now we reintroduce the drug concentration level $d \in \mathcal{D}$ and use the approximation

$$\mathbf{Y}_{d,r}(\mathcal{T}) \approx \sum_{i=1}^{K} n_i\boldsymbol{\mu}^{(i)}(\mathcal{T}, d) + N^{(r)}(0, n_i\mathbf{V}^{(i)}(\mathcal{T}, d)),$$

where $N^{(r)}(0, n_i \mathbf{V}^{(i)}(\mathcal{T}, d))$ represents independent and identical copies of a random vector with distribution $N(0, n_i \mathbf{V}^{(i)}(\mathcal{T}, d))$. We then formulate our newly proposed statistical model as

$$\mathbf{x}_{\mathcal{T},d,r} \approx \sum_{i=1}^{K} n_i \boldsymbol{\mu}^{(i)}(\mathcal{T}, d) + N^{(r)}(0, n_i \mathbf{V}^{(i)}(\mathcal{T}, d)) + N^{(r)}(0, c^2 \mathbf{I}), \tag{6}$$

where the final term captures experimental measurement error as before. Based on this statistical model, we compute the likelihood function $L(\boldsymbol{\theta}(\mathcal{K})|\mathbf{X})$, which represents the probability of obtaining the observed dataset $\mathbf{X}$ given the set $\boldsymbol{\theta}(\mathcal{K})$ of all parameters in the model:

$$\boldsymbol{\theta}(\mathcal{K}) = \{(\alpha_i, \beta_i, (\nu_{ij})_{j \neq i}, b_{i,\beta}, E_{i,\beta}, b_{i,\nu}, E_{i,\nu})_{i \in \mathcal{K}}, c\}.$$

We then apply the maximum likelihood estimation framework to obtain the parameter set that is most likely to explain the observed dataset. The estimated parameter set $\hat{\boldsymbol{\theta}}(\mathcal{K})$ is computed by minimizing the negative log-likelihood, which is equivalent to maximizing the likelihood function:

$$\hat{\boldsymbol{\theta}}(\mathcal{K}) = \underset{\boldsymbol{\theta}(\mathcal{K}) \in \hat{\boldsymbol{\Theta}}}{\arg \min} - \log(L(\boldsymbol{\theta}(\mathcal{K})|\mathbf{X})), \tag{7}$$

where $\hat{\boldsymbol{\Theta}}$ is the set of feasible parameters. In Table 1, we provide an overview of notation.

Notation	Dimension	Description	Definition/Range
α_i	1	Subtype i symmetric division rate	$\alpha > 0$
$\beta_i(d)$	1	Subtype i death rate	$\beta(0) \geq 0$
$\nu_{ij}(d)$	1	Subtype i asymmetric division rate to subtype j	$\nu(0) \geq 0$
$\lambda_1(d)$	1	*Long-run growth rate*	$\lambda_1(0) > 0$
$\mathcal{K}$	$1 \times K$	Subpopulation index set	$|\mathcal{K}| = K$
$\mathbf{B}(t, d)$	$1 \times K$	Vector of subpopulation counts	$\mathbf{B}(t, d) \geq \mathbf{0}$
$\mathbf{B}^{(i)}(t, d)$	$1 \times K$	$\mathbf{B}(t)$ generated from single type i cell	$\mathbf{B}^{(i)}(t, d) \geq \mathbf{0}$
$\mathbf{m}^{(i)}(t, d)$	$1 \times K$	Expected value of $\mathbf{B}^{(i)}(t, d)$	$\mathbf{m}^{(i)}(t, d) \geq \mathbf{0}$
$\boldsymbol{\Xi}^{(i)}(t, d)$	$K \times K$	Covariance matrix of $\mathbf{B}^{(i)}(t, d)$	$det(\boldsymbol{\Xi}^{(i)}(t, d)) > 0$
$\mathbf{A}(d)$	$K \times K$	Infinitesimal generator matrix	$\mathbf{A}_{ii}(d) = \alpha_i - \beta_i(d), \mathbf{A}_{ij}(d) = \nu_{ij}(d)$
$\mathbf{M}(t, d)$	$K \times K$	Mean matrix	$\lim_{t \to 0} \mathbf{M}(t, d) = \mathbf{I}$
$\mathbf{I}$	$K \times K$	Identity matrix	$\mathbf{I}$
$\mathbf{C}^{(i)}(t, d)$	$K \times K$	Factor matrix for covariance	see Proposition 1
b	1	Maximum drug effect parameter	$b > 0$
E	1	Half maximum drug effect parameter	$E > 0$
m	1	Drug effect steepness parameter	$n > 0$
n	1	Initial total cell count	$n \in [1000, 10000]$
c	1	Standard deviation of the i.i.d. observation noise	$c \in (0, 0.1n)$
$\mathbf{p}$	$1 \times K$	Vector of initial proportion p_i for subtype i	$\sum_{i=1}^{K} p_i = 1, n_i = np_i$
$\boldsymbol{\pi}(d)$	$1 \times K$	*Stable proportion*	$\boldsymbol{\pi}(d)\mathbf{A}(d) = \lambda_1(d)\boldsymbol{\pi}(d)$
$\mathcal{D}$	$1 \times N_D$	Set of concentration levels applied	$|\mathcal{D}| = N_D$
$\mathcal{T}$	$1 \times N_T$	Set of time points	$|\mathcal{T}| = N_T$
$\mathbf{W}_{n_i}^{(i)}(\mathcal{T}, d)$	$1 \times N_T$	Centered and normalized process of total cell counts	$\mathbf{W}_{n_i}^{(i)}(\mathcal{T}, d) = \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \left(\mathbf{y}_j^{(i)}(\mathcal{T}, d) - \boldsymbol{\mu}^{(i)}(\mathcal{T}, d) \right)$
$\boldsymbol{\mu}^{(i)}(\mathcal{T}, d)$	$1 \times N_T$	Expected value of the total cell counts	$\mu^{(i)}(\mathcal{T}, d) \in \mathbb{R}^{1 \times N_T}$
$\mathbf{V}^{(i)}(\mathcal{T}, d)$	$N_T \times N_T$	Covariance matrix of $\mathbf{W}_{n_i}^{(i)}(\mathcal{T}, d)$	$\mathbf{V}^{(i)}(\mathcal{T}, d) \in \mathbb{R}^{N_T \times N_T}$

Table 1: Table of definitions in Section 2.1, Section 2.2, Section 2.3, and Section 2.4.

2.5 Simplified model of drug effect on CSCs and CNSCs mixture

In our computational analysis of Sections 3 and 4, we will investigate the performance of our newly proposed framework on a tumor population consisting of two phenotypes: 1. CSCs, 2. CNSCs.

Our primary focus is on identifying drug-induced plasticity, that is drug-induced transitions from the CNSC phenotype to the CSC phenotype. Given that we only have two subpopulations, we identify the parameters for the CSC subpopulation with an s subscript and those for the CNSC subpopulation with a d subscript, as shown in Table 2:

Cell type	Initial proportion	α rate	β rate	ν rate	b for β	E for β	b for ν	E for ν
CSCs	p_s	α_s	β_s	ν_{sd}	$b_{s,\beta}$	$E_{s,\beta}$	$b_{s,\nu}$	$E_{s,\nu}$
CNSCs	p_d	α_d	β_d	ν_{ds}	$b_{d,\beta}$	$E_{d,\beta}$	$b_{d,\nu}$	$E_{d,\nu}$

Table 2: CSCs and CNSCs subpopulation parameters

To investigate heterogeneous drug effects on CSCs and CNSCs, we adopt assumptions inspired by the experimental work [25]. First, we specify assumptions related to the cell growth dynamics:

Assumption 1. *CNSCs proliferate with less potential than CSCs, while CSCs have a non-negative natural proliferation rate and a non-negative differentiation rate, expressed as $\kappa_s \geq \kappa_d \geq 0, \nu_{sd} \geq 0$.*

Assumption 2. *CNSCs do not exhibit natural plasticity, implying no transition from CNSCs to CSCs in the absence of drug.*

We note that under Assumption 2, the irreducibility condition of Section 2.1 is violated in the absence of drug. However our framework can still be applied under the following non-extinction assumption, as discussed further in Appendix 6.1.

Assumption 3. *CSCs do not go extinct, ensuring the persistence of the CSC subpopulation.*

Additionally, we constrain the pharmacologic dynamics of cells as follows:

Assumption 4. *The drug does not affect CSCs, reflecting a resistant behavior in CSCs.*

We note that if the starting population of CSCs is sufficiently large and CSCs have a positive net growth rate, $\kappa_s > 0$, then Assumption 3 will follow from Assumption 4, since a large starting population of proliferating cells unaffected by the drug is unlikely to go extinct.

The final assumption pertains to the initial distribution between CSCs and CNSCs. In [25], the researchers experimentally separated CSCs and CNSCs, and they performed assays on isolated subpopulations of CNSCs. However, assuming the *stable proportion* between CSCs and CNSCs at the start (Section 2.3) – without employing a separation technique – is more natural. Hence, we begin with Assumption 5 and later relax it in Section 3.2.2.

Assumption 5. *The initial proportion of CSCs and CNSCs in the absence of the drug is the stable proportion.*

By adopting this assumption, there is no need to estimate the initial proportions p_s and p_d; these parameters are inferred directly from the cell dynamic parameters (α, β, ν) for each subpopulation.

The five assumptions above are summarized in Table 3, each accompanied by its respective mathematical formulation.

3   Results (*in silico*)

To evaluate our framework's performance in analyzing *in silico* data, we utilized the Gillespie algorithm [5] to generate computer-simulated data based on the asymmetrical birth multi-type

Assumptions	Reflection in parameter
Assumption 1	$\kappa_s \geq \kappa_d \geq 0, \nu_{sd} \geq 0$
Assumption 2	$\nu_{ds} = 0$
Assumption 4	$b_{s,\beta} = b_{s,\nu} = 1$
Assumption 5	$\mathbf{p} = \boldsymbol{\pi}(0)$

Table 3: Assumptions and corresponding parameter constraints

branching process model. Each simulation started with an initial total cell count of $n = 1000$, and we conducted $N_R = 20$ replicates across various concentration levels $\mathcal{D}$ and time points $\mathcal{T}$, where

$$\mathcal{D} = \{0, 0.0313, 0.0625, 0.125, 0.25, 0.375, 0.5, 1.25, 2.5, 3.75, 5\},$$

$$\mathcal{T} = \{0, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36\}.$$

Finally, Gaussian observation noise was added to obtain the *in silico* dataset. Further details on the data generation can be found in Appendix 6.5.

In these *in silico* experiments, we derive two types of estimation results: 1. point estimation (PE) obtained for each experiment through the MLE process (see Appendix 6.5 for implementation details), 2. confidence intervals (CIs) obtained using the bootstrapping technique (see Appendix 6.6 for implementation details), specifically for selected *in silico* experiments.

Given our focus on a specialized case involving two subpopulations, as outlined in Section 2.5, we set $K = 2$. It is worth noting that one can treat K as a variable and employ model selection techniques to estimate the number of subpopulations exhibiting distinct phenotypic drug responses. Similar approaches were explored in [16] using a simpler statistical framework, but a detailed exploration of this topic is outside the scope of the current study.

3.1 Deconvolution of drug effects (base experiment)

3.1.1 Illustrative example

We begin with a single illustrative example. Table 4 shows the true parameter set used to generate the *in silico* dataset and the corresponding PEs obtained using our statistical framework. In Figure 3, we display both the PEs and CIs for this example, focusing on two key metrics: the *stable proportion* $\boldsymbol{\pi}(d)$ between phenotypes under each concentration level d, and the GR_{50} dose for the CNSC growth rate (cytotoxic effect) and dedifferentiation rate (plasticity effect), respectively. The GR_{50} dose is the concentration level at which the drug achieves half its maximal observed effect on the growth rate. It is a useful summary metric of the drug effect which incorporates both primary drug effect parameters E and b, as is further discussed in Appendix 6.7.

	α_s	β_s	ν_{sd}	α_d	β_d	$b_{d,\beta}$	$E_{d,\beta}$	$b_{d,\nu}$	$E_{d,\nu}$
$\boldsymbol{\theta}^*$	0.5407	0.5055	0.2929	0.2280	0.2280	0.8536	0.7073	1.0827	1.2285
$\hat{\boldsymbol{\theta}}$	0.5407	0.5036	0.2988	0.2412	0.2419	0.8612	0.6468	1.0773	1.2336

Table 4: True parameter set $\boldsymbol{\theta}^*$ and point estimates $\hat{\boldsymbol{\theta}}$ in Figure 3

Figure 3 illustrates that our newly proposed framework accurately estimates the *stable proportion* in the absence of drug (pie charts in top panel). Additionally, the varying *stable proportions* of CSCs at different drug concentrations are captured by the CIs constructed from the bootstrapping
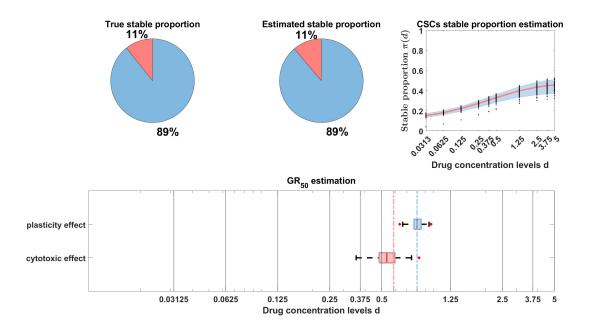
10

Figure 3: Estimation of *stable proportion* and GR_{50} based on the data generated by the true parameter set in Table 4. The pie chart demonstrates the PE of the *stable proportion* between CSCs and CNSCs under a no-drug environment. The upper-right plot shows 100 bootstrapped PEs as scatter points and the corresponding 90% CIs for the *stable proportion* of CSCs at each drug concentration level, with the red line representing the true *stable proportion* of CSCs in the same plot. The boxplot illustrates the CIs of GR_{50} of both the drug-induced plasticity effect and cytotoxic effect, with the dashed colored lines representing the corresponding true GR_{50} values.

estimations (line plot in top panel). The CIs also encompass the correct GR_{50} values for both the cytotoxic and plasticity effects (bottom panel).

In this example, the true GR_{50} values for the two drug effects fall between two concentration levels applied in the experiment. In our previous work [35], we used a similar statistical framework to investigate a tumor with two subpopulations, where each was affected by an anti-cancer drug to a varying degree and there were no transitions between the subpopulations. There, we observed that close GR_{50} values for the cytotoxic effects on each subpopulation often resulted in poor estimations of the GR_{50} values. In the current study, where we consider two distinct types of drug effects, the difference in GR_{50} is no longer critical for distinguishing these effects.

3.1.2 Estimation across a wide range of true parameter sets

To assess estimation performance across a wide range of biologically realistic parameter sets, we examined the relative errors (RE) from 100 datasets, each generated using a different true parameter set. The RE between an estimator $\hat{x}$ and the true value x^* was calculated as

$$Er(\hat{x}; x^*) = \left| \frac{\hat{x} - x^*}{x^*} \right|. \tag{8}$$

11

The results for all datasets are summarized using boxplots in Figure 4. Note that we recorded the RE for $1 - b_{d,\beta}$ and $b_{d,\nu} - 1$ rather than $b_{d,\beta}$ and $b_{d,\nu}$ directly. These new measurements of REs ensure symmetry when measuring the maximum drug effect between cytotoxic and plasticity effects, since $b_{d,\beta} \in (0.5, 1)$ and $b_{d,\nu} \in (1, 1.5)$ (see Appendix 6.5). Additionally, we included a horizontal line at RE $= 0.2$ as a threshold to indicate reasonably well estimated parameters.
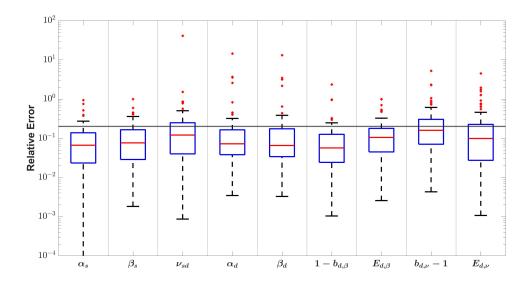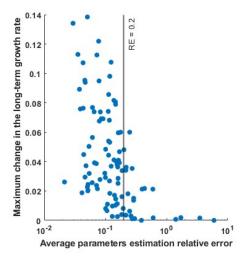
Figure 4: Relative error of the parameter estimation. The boxplot contains 100 independent experiments generated by randomly selected true parameter sets. The red line is the median of the REs. The grey horizontal line is a threshold when RE is equal to 0.2.

Figure 4 shows that most point estimates are reasonable, with over three fourths of estimates having a RE below the specified threshold. This indicates that our newly proposed framework can successfully disentangle the cell growth dynamics and drug effect dynamics. We note that the plasticity effect parameter $b_{d,\nu}$ is generally estimated less accurately than the cytotoxic effect parameter $b_{d,\beta}$. This may stem from the fact that in all parameter sets, the true plasticity effect is smaller than the cytotoxic effect, i.e. $b_{d,\nu} - 1 < 1 - b_{d,\beta}$, making it more challenging to detect the plasticity effect (see Appendix 6.5). In addition, we note that the relative error metric is more sensitive when the true value is small, since even small absolute errors $|\hat{x} - x^*|$ can lead to large relative errors $|(\hat{x} - x^*)/x^*|$.

Although our estimation framework performs well overall, there are some examples where it fails to recover the true values from the data. One explanation for low quality in the estimation is a poor experimental design, i.e. poor selection of concentration levels and data collection time points, resulting in reduced parameter identifiability. For instance, when estimating E, if the tested concentration levels do not cover the true value, i.e. $\max(\mathcal{D}) < E$, it will be challenging for our framework to identify the true value. Even when E is within the tested concentration levels, observation noise and the stochastic nature of the data generation can still distort the estimation, particularly when E is close to the maximum tested concentration level. A detailed analysis of how experimental design impacts parameter identifiability will be explored in future work.
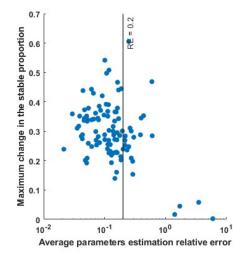
Figure 5: Scatter plots of 100 independent experiments. Both plots have average parameters estimation relative error on the x-axis. In the left panel, the y-axis represents the observable drug effect on the long-term growth rate of the CSCs and CNSCs mixture. In the right panel, the y-axis represents the observable drug effect on the *stable proportion* between the CSCs and CNSCs.

3.1.3 Analysis of failed estimations

As previously mentioned, one potential explanation for poor estimation under our framework is a poor experimental design. To better understand potential challenges in parameter identifiability, we analyzed the true parameter sets and the simulated data from the 100 experiments described earlier. We specifically examined the drug effects on two theoretical metrics of long-run behavior: *stable proportion* ($\boldsymbol{\pi}(d)$) and *long-run growth rate* ($\lambda_1(d)$) (see Section 2.3). For each metric, we computed its theoretical values across different concentration levels and determined the maximum discrepancy among them to quantify the ranges of drug effects observed in the experiments.

In Figure 5, scatter plots illustrate these quantities plotted against the average RE over all estimated parameters for each experiment. In the right panel's lower right corner, four examples show a maximum change in stable proportion below 0.1, coinciding with a notably high average estimation error. In the left panel, these same four examples exhibit low maximum changes in long-run growth rates. This observation indicates that inaccurate estimations tend to occur when observed drug effects are minimal. This insight is further supported by Figure 6, where we visualized two example datasets with high vs. low RE, and observed that the dataset with less variation in growth pattern across different concentration levels has higher average RE. To summarize, poor estimation in terms of RE can in most cases be traced to a small observed drug effect, which can either arise due to a poor experimental design (drug effects do not manifest due to poor selection of concentration levels or data collection points) or a small true drug effect, in which case the RE metric is more sensitive to estimation inaccuracy as mentioned in the previous section.

3.2 Robustness of the framework

To assess the robustness of our newly proposed statistical framework using *in silico* data, we designed a series of experiments to evaluate the performance of the framework when relaxing some
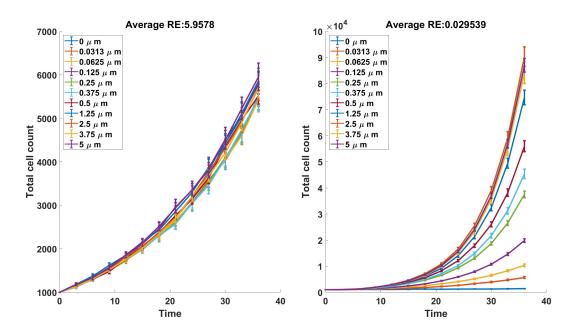
13

Figure 6: Visualization of the simulated data with sufficient and insufficient information for accurate estimation. The left panel demonstrates data with insufficient information, while the right panel demonstrates data with sufficient information. Each line plot illustrates total cell count data across 13 time points under a specific concentration level described in the legend. The error bar is derived from the standard deviation of 20 independent replicates.
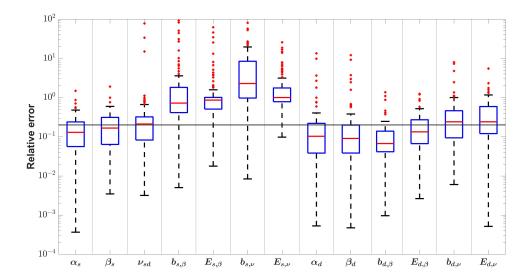
14

Figure 7: Relative error of the parameter estimation when relaxing Assumption 4. The remaining details of the plot can be found in Figure 4.

of the assumptions in Section 2.5. Detailed experimental settings are described in Appendix 6.5.

3.2.1 Relaxed drug effect assumption

In Assumption 4 of Section 2.5, we posit that the drug will not affect the CSCs. However, this assertion may seem too stringent for practical scenarios, where the varying microenvironment caused by different drug concentration levels could potentially affect the CSCs [37]. Therefore, we conduct another set of experiments to examine the performance of our framework when the drug does influence the CSCs. While the mechanisms underlying drug resistance in CSCs are not yet fully understood [28, 19], for testing purposes, we make the simplifying assumption that the drug affects CSCs in a similar way to CNSCs, manifesting through cytotoxic effects and increased asymmetrical divisions, albeit to a lesser extent than for CNSCs. As noted in Section 2.2, our framework is also adaptable to scenarios where the drug might decrease the differentiation rate.

Figure 7 presents the RE from 100 independent experiments. We observe that the estimates of certain parameters, such as $(\nu_{sd}, b_{d,\nu}, E_{d,\nu})$, show deterioration compared to Figure 4, with more than 50% of estimates having RE above the 0.2 threshold value. However, our framework continues to provide reasonably accurate estimates for parameters like $\alpha_d, \beta_d, b_{d,\beta}$.

As mentioned in the previous section, the estimation accuracy may be correlated to the significance of the drug effect. In the current experiment, we find more compelling evidence to support this observation. In particular, the estimation accuracy for the CSC drug effect parameters $(b_{s,\beta}, E_{s,\beta}, b_{s,\nu}, E_{s,\nu})$ is significantly lower than the accuracy in estimating the analogous parameters for CNSCs, due to our assumption of a more pronounced impact of the drug on CNSCs.
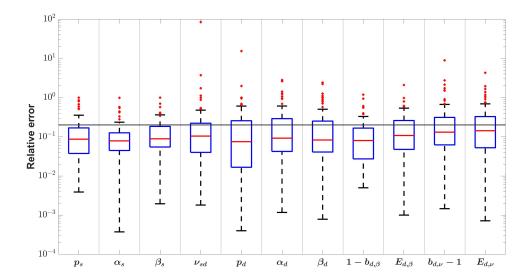
15

Figure 8: Relative error of the parameter estimation when relaxing Assumption 5. The remaining details of the plot can be found in Figure 4.
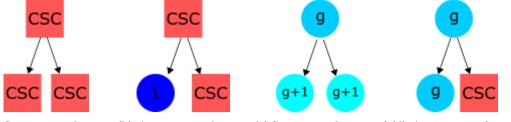
3.2.2 Relaxed initial proportion assumption

Our next experiment aims to investigate the possibility of estimating the initial subpopulation structure under arbitrary initial proportions, relaxing Assumption 5 of Section 2.5. To accommodate this situation, we reintroduce two parameters, p_s and p_d, with the constraint that $p_s + p_d = 1$, representing the initial proportions of CSCs and CNSCs respectively.

The results in Figure 8 indicate that relaxing the initial proportion assumption does not deteriorate the accuracy of the estimation of other parameters, and the initial proportion is estimated reasonably well. This highlights the potential of our framework to deconvolute the phenotypic subpopulation structure with an unknown initial proportion. It is also worth noting that we allow the p_s and p_d to be uneven, for example, $p_s = 0.03$ and $p_d = 0.97$.

3.2.3 Limited division CNSC dynamics assumption

We finally consider a more general CNSC division scheme, where CNSCs can only undergo symmetric division a limited number of times, reflecting a gradual loss of proliferation potential. We refer to CNSCs produced through asymmetric division of CSCs as first-generation CNSCs, which can only divide up to G times. For the first G generations of CNSCs, we assume that the g-th generation can symmetrically divide into two $(g + 1)$-th generation CNSCs, asymmetrically divide into one CSC and one g-th generation CNSC, or die. The $(G + 1)$-th generation CNSCs can only die or asymmetrically divide into one CSC and one $(G + 1)$-th generation CNSC. All CNSCs share the same symmetric division rate α_d, death rate β_d and asymmetric division rate ν_{ds} (Figure 9). In addition, the drug effect parameters $b_{d,\beta}, E_{d,\beta}, b_{d,\nu}, E_{d,\nu}$ are the same for CNSCs in the different generations. This scheme encompasses a broader spectrum of CNSC dynamics, ranging from the scenario where CNSCs cannot self-renew ($G = 0$) to the scenario where CNSCs have an infinite capacity for self-renewal with a certain rate α_d ($G \to \infty$). Theoretically, our current framework can

accurately capture the extreme scenarios where $G = 0$ or $G \to \infty$. We are interested in the practical performance of our framework in the intermediate case, i.e., $0 < G < \infty$.

(a) Symmetric division of CSC (rate: α_s)

(b) Asymmetric division of CSC (rate: ν_{sd})

(c) Symmetric division of CNSC (rate: α_d)

(d) Asymmetric division of CNSC (rate: ν_{ds})

Figure 9: Illustration of four distinct division events assumed for the limited division scenario. Square shape represents CSC and round shape indicates CNSC, while different colors illustrate a different division potential of each generation in CNSC.

We first want to determine whether our framework can reliably estimate the symmetric division rate of CNSCs, α_d. To streamline the experiment, we selected the same true value $\alpha_d^* > 0$ for $G \in \{0, 1, 2, 3\}$. Instead of using the RE, we adopt the ratio between the estimated value and the true value of α_d as the metric of accuracy. Specifically, for an estimator $\hat{x}$ and the true value x^*, we compute the accuracy metric as:

$$Er(\hat{x}; x^*) = \frac{\hat{x}}{x^*}. \tag{9}$$

This new metric effectively captures two important scenarios: (i) the framework should produce an estimation $\hat{\alpha}_d = 0$ when $G = 0$, resulting in a ratio $\hat{\alpha}_d / \alpha_d^* = 0$, and (ii) the framework should accurately estimate $\hat{\alpha}_d = \alpha_d^*$ when G is sufficiently large.

The results of our experiment (Figure 10) align well with these two scenarios. Specifically, our framework accurately infers that $\alpha_d = 0$ when the maximum division number is $G = 0$, indicating that CNSCs cannot undergo self-renewal. However, when $G = 1$, 75% of estimates for α_d are below the true value. Surprisingly, our framework accurately recovers the true value of α_d when $G = 2$ and $G = 3$. For other parameters, we generally observed systematic biases in the estimation, see Appendix 6.8. However, interestingly, we found that the drug effects' inflection points, $E_{d,\nu}$ and $E_{d,\beta}$, were accurately estimated across different values of G. This observation demonstrates the robustness of our framework in estimating the drug effects' inflection points under varying differentiation dynamics. In other words, through detecting the changes in cell population growth dynamics, our framework can estimate the concentration levels at which such changes occur, even though it may not fully explain these changes. For the other parameters, we note that it is possible to fully capture the dynamics of CNSCs with limited proliferation potential by expanding our model beyond two subpopulations, as we further discuss in the conclusion section (Section 5).

4 Results (*in vitro*)

4.1 DATA overview

To validate our framework using *in vitro* data, we utilized data from the recent study [25] that investigated four human gastric carcinoma cell lines (AGS SORE6−, AGS SORE6+, Kato
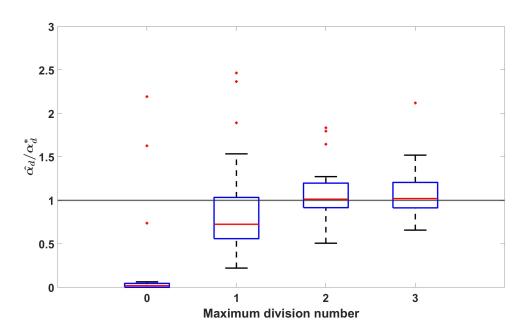
Figure 10: Estimation error of the CNSCs division rate when CNSCs can only divide a limited amount of times. The ratios between the estimated division rate and the true division rate are presented to varying maximum division numbers G. The boxplot is based on 30 independent experiments for each $G \in \{0, 1, 2, 3\}$ with varying true parameter sets. The grey horizontal line illustrates the ratio when CNSCs division rate is accurately estimated.

III SORE6−, Kato III SORE6+). Our focus in this section is specifically on the AGS SORE6−
and AGS SORE6+ cell lines. SORE6 serves as a reporter system indicating the stemness of these
cancerous gastric cells, where SORE6− and SORE6+ denote gastric CNSCs and CSCs, respectively.
In the study [25], ciclopirox olamine (CPX-O) was found to reprogram gastric CNSCs into CSCs.
To support these findings, the authors cultivated pure AGS SORE6− cells and monitored both cell
viability data and the percentage of AGS SORE6+ cells using the SORE6 reporter system.

In our study, we mainly utilized two experimental datasets provided in [25]:

1. *AGS Conc* (X_{Conc}): AGS SORE6− samples were cultivated under 9 concentration levels of
   CPX-O:
   $$\mathcal{D}_1 = \{0, 0.125, 0.25, 0.5, 1, 2, 4, 8, 16\} \; (\mu m).$$

   Total cell count data (TC) and stem cell proportion (SC) were observed after 48 hours of
   treatment, i.e. $\mathcal{T}_1 = \{48\}$ (hours).

2. *AGS Time* (X_{Time}): AGS SORE6− samples were cultivated under 3 concentration levels of
   CPX-O:
   $$\mathcal{D}_2 = \{0, 4, 8\} \; (\mu m).$$

   Total cell count data (TC) and stem cell proportion (SC) were observed after 2, 6, 12, 24 and
   48 hours of treatment:
   $$\mathcal{T}_2 = \{2, 6, 12, 24, 48\} \; (\text{hours}).$$

Unfortunately, neither of these experiments provides enough information to estimate all param-
eters under our framework. Specifically, three concentration levels in the *AGS Time* data are not
sufficient to estimate the four parameter effects of the drug in our framework, and a single time point
in the *AGS Conc* data is inadequate to capture the heterogeneous dynamics of cell growth. This
limitation may arise from the need to employ the FACS technique to distinguish between the CSC
and CNSC populations. To overcome this limitation, we propose to employ a data imputation on the
AGS Conc data based on the information from the *AGS Time* data. Further details of imputation
can be found in Appendix 6.9.2. We provide the imputed dataset in Table 5.

Concentration levels (μm)	0	0.125	0.25	0.5	1	2	4	8	16
TC of replicate 1 at time 12	3083	3083	3083	3083	3083	3083	3083	3083	3083
TC of replicate 1 at time 24	4272	4246	4222	4173	4080	3907	3608	3150	2568
TC of replicate 1 at time 48	5038	4503	3242	2235	1849	1608	1140	1054	756
TC of replicate 2 at time 12	3502	3502	3502	3502	3502	3502	3502	3502	3502
TC of replicate 2 at time 24	4851	4823	4795	4739	4633	4437	4098	3578	2917
TC of replicate 2 at time 48	5722	4597	3433	2881	2511	2295	1875	1309	694

Table 5: *In vitro* total cell count (TC) data: *AGS Conc* data with augmented estimation according
to *AGS Time*. The colored data are estimated data according to the *AGS Time* data, while the
black data are actual data from [25].

4.2 Using model selection criteria to verify the existence of drug-induced plasticity

One potential application of our newly proposed statistical framework in practical settings is to
detect the presence of drug-induced plasticity using model selection criteria. Here, we employ the

Model	$|\theta|$	Free parameters
I	12	$\alpha_s, \beta_s, \nu_{sd}, \alpha_d, \beta_d, b_{d,\beta}, E_{d,\beta}, m_{d,\beta}, b_{d,\nu}, E_{d,\nu}, m_{d,\nu}, c$
Ia	6	$\alpha_d, \beta_d, b_{d,\beta}, E_{d,\beta}, m_{d,\beta}, c$
II	14	$\alpha_s, \beta_s, \nu_{sd}, \alpha_d, \beta_d, b_{d,\beta}, E_{d,\beta}, m_{d,\beta}, b_{d,\nu}, E_{d,\nu}, m_{d,\nu}, k, t_0, c$
IIa	8	$\alpha_d, \beta_d, b_{d,\beta}, E_{d,\beta}, m_{d,\beta}, k, t_0, c$

Table 6: Free parameters in models proposed in *in vitro* result.

	Drug-induced plasticity	no Drug-induced plasticity
Models without time-delay assumption	**AIC(I): 653.4866**	AIC(Ia): 677.9985
Models with time-delay assumption	**AIC(II): 534.1237**	AIC(IIa): 556.4027

Table 7: AIC value for 4 varying model assumptions

well-known Akaike Information Criterion (AIC) calculated by the formula:

$$AIC = 2|\theta| - 2\log(L^*), \tag{10}$$

where $|\theta|$ is the number of free parameters in the model and L^* is the maximum value of the likelihood function for the model. AIC, rooted in information theory, quantifies the information lost by a model, with lower AIC values indicating higher model quality.

We consider a total of four different models. The first two models, based on the framework described in equation (6), differ in their assumptions regarding the induction of plasticity by CPX-O. We denote the model that assumes drug-induced plasticity as model I and the alternative as model Ia. In fitting the *in vitro* data, we reintroduce the Hill parameter m to our drug-effect model (Section 2.2). Additionally, it is assumed that the drug does not affect the dynamics of CSCs.

Model I has 12 free parameters, as detailed in Table 6, to depict the cell growth dynamics of CSCs and CNSCs and the drug response of CNSCs. In contrast, Model Ia requires only 6 free parameters, as it assumes no drug-induced plasticity for CPX-O, resulting in no parameters necessary for gastric CSCs. Indeed, since the *in vitro* experiments in question are started with isolated CNSC subpopulations ($p_d = 1$), and no natural plasticity of CNSCs is assumed ($\nu_{ds} = 0$), no CSCs should emerge in the experiments according to the assumptions of Model Ia.

In the other two models, we introduce a time-delayed drug-effect which is present in the *AGS Time* data, since similar cell growth patterns are observed within the first 12 hours across different drug concentration levels, with pronounced variations emerging thereafter. Further details can be found in Appendix 6.9.1. Therefore, we assume that the maximum drug effect parameter b varies over time following a two-parameter logistic function as described in Section 2.2.2. Like the previous two models, one of these time-delayed drug-effect models, denoted as II, assumes that CPX-O can induce plasticity, while the other, denoted as IIa, does not. Model II has 14 free parameters, whereas Model IIa involves 8 free parameters, as detailed in Table 6.

The AIC results, summarized in Table 7, indicate a preference for the models assuming drug-induced plasticity under the AIC model selection criteria. Furthermore, comparing the AIC values between models with and without the time-delay assumption reveals the significant role of the time-delay effect in the current data. While our newly proposed framework may not fully capture the experimental data in [25], it nonetheless yields conclusions consistent with the actual observations reported in [25]. It should also be stressed that our framework is able to infer the presence of drug-induced resistance from total cell count data, whereas this conclusion is drawn in [25] using additional data on the CSC proportion over time. Of course, the inference of drug-induced plasticity is in this

20

case facilitated by the ability to start the experiments from isolated CNSCs, but our framework is applicable to a broader range of experimental setups without this ability.

In Appendix 6.10, we validate our framework's estimations using the imatinib-sensitive and -resistant $Ba/F3$ cells dataset provided in [16], which originates from a mixture of two phenotypes with a known initial proportion. Our framework suggests that drug-induced phenotypic transitions have minimal impact on the dynamics of this dataset, consistent with our prior belief that these transitions are infrequent. Unlike the conclusion drawn in the current section, the experiment in Appendix 6.10 demonstrates that our framework can reliably avoid indicating false effects when none are present.

5   Conclusion

In this study, we introduced a novel statistical framework designed for analyzing HTS bulk data involving multiple subpopulations of cells. Using an asymmetrical birth multi-type branching process, we extended a model introduced in our previous work [35] to accommodate transitions between the different subpopulations, thereby enhancing its range of applications. One such application involves inferring drug-induced plasticity in a mixture population of cancer stem-like cells (CSCs) and cancer non-stem-like cells (CNSCs). The asymmetrical birth multi-type branching process naturally accounts for the differentiation of CSCs and the dedifferentiation of CNSCs. Additionally, we incorporated the Hill equation to characterize the cytotoxic effects of anti-cancer drugs and drug-induced dedifferentiation rates. We tested our approach using both *in silico* and *in vitro* data.

In our *in silico* experiments, we used stochastic simulation to generate datasets involving drug-treated mixtures of CSCs and CNSCs. The datasets consisted of total cell counts collected at predetermined drug concentration levels and time points. Drawing inspiration from recent research on drug-induced plasticity [25], we formulated four assumptions to simulate the dynamics of CSCs and CNSCs, along with their respective drug effects. Under these assumptions, our newly proposed framework not only identifies the presence of drug-induced plasticity, but also accurately predicts several key features of the mixture dynamics. Specifically, in an illustrative example, the framework accurately recovered the drug-affected *stable proportions* between CSCs and CNSCs across various drug concentration levels. Furthermore, it determined the GR_{50} values for both the drug's cytotoxic effect on CNSCs and its induced plasticity effect. Through further numerical experiments, we confirmed that the newly proposed framework consistently provides precise estimations for each model parameter, capturing the growth dynamics of individual subpopulations, transitions between the subpopulations, as well as drug-induced toxicity and plasticity.

We further investigated the robustness of our framework by relaxing assumptions regarding the dynamics of the CSCs and CNSCs and how the drug affects each supbopulation. By allowing the CSC and CNSC mixture to start from an arbitrary initial proportion instead of the stable drug-free proportion, our framework accurately estimated not only the mixture's growth dynamics and drug effects but also the initial proportion. However, relaxing the assumption that the drug only affects CNSCs resulted in a more complex drug effect profile, which degraded the identifiability of each individual effect. In particular, the drug effects on CSCs were less identifiable, which may be due to our assumption that the drug has a more significant effect on CNSCs. It should also be kept in mind that we only assume data on total cell counts, as opposed to cell counts for individual subpopulations, which is an inherent limitation for teasing apart complex evolutionary dynamics and drug effect profiles involving multiple parameters. Using this kind of data, some simplifying assumptions must be made to ensure the identifiability of all model parameters.

Additionally, we explored the performance of our framework under the assumption of limited

division of CNSCs, portraying the differentiation process as a gradual loss of cell proliferation potential. Our framework successfully estimated the CNSC division rate when the maximum division number (G) of CNSCs equaled 0, 2 or 3. Furthermore, it consistently provided reliable estimates of the half-maximum effects parameter (E) for each drug effect on the CNSCs. Even though parameter estimates generally degraded, these results demonstrate that our newly proposed framework can offer valuable insights even under more complex and realistic scenarios. To fully capture these dynamics, we propose expanding our model beyond two subpopulations, by modeling each generation of CNSCs as a distinct subpopulation. Specifically, we can represent the symmetric division of g-th generation cells into two $(g+1)$-th generation cells as birth events where a cell of one type produces two cells of another type. Currently, our asymmetric division framework does not account for this behavior, as it models subpopulation transitions through asymmetric divisions. However, modeling subpopulation transitions through symmetric division is straightforward using the multi-type branching process framework [1]. We plan to explore this in future work.

We proceeded to validate our framework using data from mixtures of gastric CNSCs and CSCs treated with ciclopirox olamine [25], as well as data from an imatinib-sensitive and -resistant *Ba/F3* mixture [16]. Our framework successfully demonstrated the absence of drug-induced transition in the latter dataset, which is consistent with the original experimental observations. However, validation using data from [25] posed a challenge due to the limited amount of *in vitro* data available, which was not comparable to the extensive data generated in our *in silico* experiments. This is at least in part due to the experimental procedure described in [25], which involves the use of FASC to separate CSCs and CNSCs – a step that our framework does not require. Nevertheless, our newly proposed framework indicates that CPX-O induces plasticity in the AGS gastric carcinoma cell line using the AIC model selection technique. This finding aligns with the conclusion drawn in [25].

In general, our framework requires many parameters to model various drug effects, necessitating multiple data points at distinct concentration levels to adequately capture mixture dynamics under drug influence. However, this requirement may be offset by the advantage of not needing a more sophisticated technique for subpopulation separation. Moreover, the quality of the data is equally as important as the quantity of it. As observed in our *in silico* experiments, some true parameter sets became unidentifiable under the specific concentration levels used in the experiments. To enhance the framework's utility, one future direction involves developing techniques for experimental design, meaning optimal selection of drug concentrations and time points for accurate parameter inference. Such analysis could markedly improve the framework's ability to extract insights from HTS data.

Another challenge identified in the *in vitro* experiment involving gastric CSCs and CNSCs was the time-inhomogeneous drug effect. While our drug-effect base model assumes time-homogeneous drug effects on the target tumor sample, data from [25] show significant heterogeneity in drug effects over time. In Section 2.2.2, we addressed this by developing a two-parameter logistic function, resulting in a better fit to the *in vitro* data.

There are several avenues for extension of the framework beyond those already mentioned. First, while we have focused on inferring drug-induced plasticity in a CSC and CNSC mixture, the newly proposed framework is based on a versatile mathematical model (multi-type branching process) which can be adapted to a broader range of transition dynamics between two or more subpopulations. These subpopulations are defined by their heterogeneous responses to controllable interferences – specifically, drugs in this case. Second, we have in this work focused on cellular responses to single-drug interventions. One potential extension involves expanding our framework to encompass more complex multi-drug responses, potentially unveiling more intricate underlying systems. Moreover, external factors such as nutrient levels, hypoxia, stromal content, and other microenvironmental factors could be integrated into our framework in future work.

Acknowledgements

Code availability

The code necessary to reproduce the experimental results presented in this paper are published in `https://github.com/chenyuwu233/Cancer-Stem-Cells-Drug-induced-Plasticity`.

6 Appendix

6.1 Reducible infinitesimal generator matrix A and long-run behavior of B

The results in Section 2.3 rely on the irreducibility assumption of infinitesimal generator matrix $\mathbf{A}$. However, Assumption 2 in Section 2.5 violates this irreducibility assumption under no-drug conditions. Consequently, we explore the details of these assumptions and discuss the situation when those assumptions are violated in this section.

The main issue arising from reducibility is that the process $\mathbf{B}(t, 0)$ may not converge almost surely to a *stable proportion* $\boldsymbol{\pi}$ in the long-run. To address this, we use the results in [23]. Given that we assume $\kappa_s > \kappa_d > 0$ (Assumption 1) and CSCs do not go extinct (Assumption 3), we can directly apply Proposition 1 and Corollary 1 in [23]. Specifically, Proposition 1 in [23] states that there exist random variables V_s and V_d such that

$$\lim_{t \to \infty} e^{-\kappa_s t} B_s(t) = V_s \text{ and } \lim_{t \to \infty} e^{-\kappa_s t} B_d(t) = V_d$$

almost surely. From Corollary 1 in [23], we further have

$$V_d = \frac{\nu_{sd} V_s}{\kappa_s - \kappa_d}$$

almost surely. From this ratio, we can further derive

$$[V_s, V_d] = C \left[\frac{\kappa_s - \kappa_d}{\kappa_s - \kappa_d + \nu_{sd}}, \frac{\nu_{sd}}{\kappa_s - \kappa_d + \nu_{sd}} \right] = C\boldsymbol{\pi},$$

where C is a random variable and the *stable proportion* $\boldsymbol{\pi}$ is still the left eigenvector of $\mathbf{A}$ corresponding to the largest eigenvalue. Consequently, the *stable proportion* exists. By following the same procedure—obtaining the left eigenvector of $\mathbf{A}$ corresponding to the largest eigenvalue—we can determine the *stable proportion* when $\mathbf{A}$ becomes reducible under the no-drug condition.

6.2 Long-run behavior of finite moment of the multi-type branching process

In [1], the authors reviewed the *long-run growth rate* of the first and the second moments of the multi-type branching process. Specifically, the *mean matrix* scaled by $e^{-\lambda_1 t}$ would converge to a

matrix independent of time. In other words, the long-run expected cell count will increase at rate $e^{\lambda_1 t}$. Similarly, the second moment has a growth rate $e^{2\lambda_1 t}$. To prove Proposition 2, we extend these results by considering any finite moment of this process, i.e. we want to show that for any positive finite integer z and combination $\mathbf{c}_z : [c_1(z), \cdots, c_K(z)] \geq 0$ such that $\sum_{i=1}^{K} c_i(z) = z$,

$$\limsup_{t \to \infty} \mathrm{E}\Big[B_1^{(i)}(t)^{c_{1(z)}} B_2^{(i)}(t)^{c_{2(z)}} \cdots B_K^{(i)}(t)^{c_{K(z)}} \Big] e^{-\lambda_1 z t} < \infty.$$

Consequently, we show the following Lemma 1.

Lemma 1. *For any integer $z > 0$ and combination $\mathbf{c}_z$, we have*

$$\mathrm{E}\Big[B_1^{(i)}(t)^{c_{1(z)}} B_2^{(i)}(t)^{c_{2(z)}} \cdots B_K^{(i)}(t)^{c_{K(z)}} \Big] = O(e^{\lambda_1 z t}).$$

Proof. The probability-generating function for the multi-type branching process for $\mathbf{B}^{(i)}(t)$ is denoted as

$$F^{(i)}(\mathbf{s}, t) := \mathrm{E}\Big[\mathbf{s}^{\mathbf{B}^{(i)}(t)} \Big] = \mathrm{E}\Big[\prod_{k=1}^{K} s_k^{\mathbf{B}_k^{(i)}(t)} \Big],$$

where $\mathbf{s} = [s_1, \cdots, s_K]$. With these notations, we can write the Kolmogorov forward equation [1] for $\mathbf{B}^{(i)}(t)$ as

$$\frac{\partial}{\partial t} F^{(i)}(\mathbf{s}, t) = \sum_{k=1}^{K} u^{(k)}(\mathbf{s}) \frac{\partial}{\partial s_k} F^{(i)}(\mathbf{s}, t), \tag{11}$$

where the function $u^{(i)}(s)$ for type i is defined as

$$u^{(i)}(\mathbf{s}) := \alpha_i s_i^2 + \beta_i + \sum_{j \neq i} \nu_{ij} s_i s_j - \Big(\alpha_i + \beta_i + \sum_{j \neq i} \nu_{ij} \Big) s_i, \quad 0 \leq \mathbf{s} \leq 1$$

We denote a vector $\mathbf{c}_z = [c_{1(z)}, c_{2(z)}, \cdots, c_{K(z)}]$ as the combination index vector. Then we use $\partial \mathbf{s}^{\mathbf{c}_z}$ to represent $\partial s_1^{c_{1(z)}} \partial s_2^{c_{2(z)}} \cdots \partial s_K^{c_{K(z)}}$.

Now we can define

$$\ell_z(t) = \max_{\mathbf{c}_z} \frac{\partial^z}{\partial \mathbf{s}^{\mathbf{c}_z}} F^{(i)}(\mathbf{s}, t) \Big|_{\mathbf{s}=1}, \tag{12}$$

and we denote $\mathbf{c}_z^*$ as the maximum combination index of z-th moment. We can see from the equation (12) that for each $z \geq 1$ there exists $C_z' > 0$ such that for any combination $\mathbf{c}_z$

$$\ell_z(t) \geq \frac{\partial^z}{\partial \mathbf{s}^{\mathbf{c}_z}} F^{(i)}(\mathbf{s}, t) \Big|_{\mathbf{s}=1}$$

$$\geq C_z' \sum_{n_1=c_{1(z)}, \cdots, n_K=c_{K(z)}}^{\infty} n_1^{c_{1(z)}} \cdots n_K^{c_{K(z)}} P(B_1^{(i)}(t) = n_1, \cdots, B_K^{(i)}(t) = n_K)$$

$$= C_z' \mathrm{E}\Big[\prod_{k=1}^{K} B_k^{(i)}(t)^{c_{k(z)}} \Big] - C_z' \mathrm{E}\Big[\prod_{k=1}^{K} B_k^{(i)}(t)^{c_{k(z)}}; B_k^{(i)}(t) < c_{k(z)} \Big].$$

As the multi-type branching process goes to infinity or extinct with probability 1 as time goes to infinity, we have, in the supercritical case,

$$\lim_{t \to \infty} \frac{\mathrm{E}\Big[\prod_{k=1}^{K} B_k^{(i)}(t)^{c_{k(z)}^*}; B_k^{(i)}(t) < c_{k(z)}^* \Big]}{\mathrm{E}\Big[\prod_{k=1}^{K} B_k^{(i)}(t)^{c_{k(z)}^*} \Big]} = 0$$

24

so that there exists C_z such that $\ell_z(t) \geq C_z \operatorname{E}\left[\prod_{k=1}^{K} B_k^{(i)}(t)^{c_{k(z)}^*}\right]$.

Thus, it suffices to establish the upper bound on $\ell_z(t)$ to prove the Lemma. We show the upper bound of $\ell_z(t)$ by induction on z. The base case was given by the fact that $e^{\lambda_1 t}$ is the largest eigenvalue of the *mean matrix* $\mathbf{M}(t)$, i.e.

$$\operatorname{E}\left[B_k^{(i)}(t)\right] = [\mathbf{e}_i \mathbf{M}(t)]_k \leq e^{\lambda_1 t},$$

where $[\cdot]_k$ is the k-th element of the vector, for any $k \in \{1, \cdots, K\}$. Then we assume that for $z > 1$ and all $j \leq z - 1, \ell_j(t) = O(e^{\lambda_1 j t})$. Then from the equation (11) we have that

$$
\begin{aligned}
\frac{d}{dt}\ell_z(t) &= \frac{\partial^z}{\partial \mathbf{s}^{\mathbf{c}_z^*}} \sum_{k=1}^{K} u^{(k)}(\mathbf{s}) \frac{\partial}{\partial s_k} F(\mathbf{s}, t)\Big|_{\mathbf{s}=\mathbf{1}} \\
&= \sum_{j=0}^{z} \sum_{\mathbf{c}_j \leq \mathbf{c}_z^*} \binom{c_{1(z)}^*}{c_{1(j)}} \cdots \binom{c_{K(z)}^*}{c_{K(j)}} \sum_{k=1}^{K} \frac{\partial^j}{\partial \mathbf{s}^{\mathbf{c}_j}} u^{(k)}(\mathbf{s}) \frac{\partial^{z-j+1}}{\partial \mathbf{s}^{\mathbf{c}_z^* - \mathbf{c}_j} \partial s_k} F(\mathbf{s}, t)\Big|_{\mathbf{s}=\mathbf{1}} \\
&= \sum_{j=1}^{z} \sum_{\mathbf{c}_j \leq \mathbf{c}_z^*} \binom{c_{1(z)}^*}{c_{1(j)}} \cdots \binom{c_{K(z)}^*}{c_{K(j)}} \sum_{k=1}^{K} \frac{\partial^j}{\partial \mathbf{s}^{\mathbf{c}_j}} u^{(k)}(\mathbf{s}) \frac{\partial^{z-j+1}}{\partial \mathbf{s}^{\mathbf{c}_z^* - \mathbf{c}_j} \partial s_k} F(\mathbf{s}, t)\Big|_{\mathbf{s}=\mathbf{1}} \\
&\leq \sum_{j=1}^{z} \sum_{\mathbf{c}_j \leq \mathbf{c}_z^*} \binom{c_{1(z)}^*}{c_{1(j)}} \cdots \binom{c_{K(z)}^*}{c_{K(j)}} \sum_{k=1}^{K} \frac{\partial^j}{\partial \mathbf{s}^{\mathbf{c}_j}} u^{(k)}(\mathbf{1}) \ell_{z-j+1}(t) \\
&\leq \lambda_1 z \ell_z(t) + \sum_{j=2}^{z} \sum_{\mathbf{c}_j \leq \mathbf{c}_z^*} \binom{c_{1(z)}^*}{c_{1(j)}} \cdots \binom{c_{K(z)}^*}{c_{K(j)}} \sum_{k=1}^{K} \frac{\partial^j}{\partial \mathbf{s}^{\mathbf{c}_j}} u^{(k)}(\mathbf{1}) \ell_{z-j+1}(t)
\end{aligned}
$$

The third equality is due to the fact that $u^{(k)}(\mathbf{1}) = 0$ for all k, the first inequality follows from the definition of $\ell_z(t)$, and the last inequality is due to the following inequality

$$\sum_{k=1}^{K} \frac{\partial}{\partial s_i} u^{(k)}(\mathbf{1}) = \alpha_i - \beta_i + \sum_{k \neq i} \nu_{ki} = \mathbf{1}\mathbf{A}\mathbf{e}_i^\intercal \leq \lambda_1,$$

for any $i \in \{1, \cdots, K\}$. Since $\ell_{z-j+1}(t) = O(e^{\lambda_1 (z-j+1)t})$, for $z \geq j \geq 2$, there exists constants $\alpha_{z, \mathbf{c}_j}$ such that

$$
\begin{aligned}
e^{-\lambda_1 z t} \ell_z(t) &\leq \sum_{j=2}^{z} \sum_{\mathbf{c}_j \leq \mathbf{c}_z^*} \binom{c_{1(z)}^*}{c_{1(j)}} \cdots \binom{c_{K(z)}^*}{c_{K(j)}} \sum_{k=1}^{K} \alpha_{z, \mathbf{c}_j} \int_0^t e^{-\lambda_1 \tau (j-1)} d\tau \\
&= \sum_{j=2}^{z} \sum_{\mathbf{c}_j \leq \mathbf{c}_z^*} \binom{c_{1(z)}^*}{c_{1(j)}} \cdots \binom{c_{K(z)}^*}{c_{K(j)}} \sum_{k=1}^{K} \frac{\alpha_{z, \mathbf{c}_j}}{\lambda_1 (j-1)} \left(1 - e^{-\lambda_1 t(j-1)}\right).
\end{aligned}
$$

The induction hypothesis is thus established, completing the proof. $\qquad \square$

6.3 Derivation of the covariance for asymmetric birth model:

In this sub-section, we would like to derive the covariance matrix

$$\boldsymbol{\Xi}^{(i)}(t) = \operatorname{E}\left[\left(\mathbf{B}^{(i)}(t) - \mathbf{m}^{(i)}(t)\right)^\intercal \left(\mathbf{B}^{(i)}(t) - \mathbf{m}^{(i)}(t)\right)\right].$$

25

Rather than drive this directly, we denote a second factorial moment $\mathbf{D}_{jk}^{(i)}(t) = \mathrm{E}\Big[\mathbf{B}_j^{(i)}(t)(\mathbf{B}_k^{(i)}(t) - \delta_{jk})\Big]$, where δ_{jk} is the Kronecker delta. Note that

$$\mathbf{D}^{(i)}(t) = \boldsymbol{\Xi}^{(i)}(t) + \mathbf{m}^{(i)}(t)^\intercal \mathbf{m}^{(i)}(t) - \mathrm{diag}(\mathbf{m}^{(i)}(t)) \tag{13}$$

From the proof of Lemma 1, we have the probability generating function for $B^{(i)}(t)$ defined as

$$F^{(i)}(\mathbf{s}, t) := \mathrm{E}\Big[\mathbf{s}^{B^{(i)}(t)}\Big] = \mathrm{E}\Bigg[\prod_{k=1}^{K} s_k^{B_k^{(i)}(t)}\Bigg],$$

where $\mathbf{s} = [s_1, \cdots, s_K]$. Then from the Kolmogorov Forward Equation (11), we have the $u^{(i)}(\mathbf{s})$ for asymmetric birth model defined as:

$$u^{(i)}(\mathbf{s}) := \alpha_i s_i^2 + \beta_i + \sum_{j \neq i} \nu_{ij} s_i s_j - \Bigg(\alpha_i + \beta_i + \sum_{j \neq i} \nu_{ij}\Bigg) s_i, \quad 0 \leq \mathbf{s} \leq 1. \tag{14}$$

Then we have

$$\frac{\partial}{\partial t}\left(\frac{\partial^2}{\partial s_j \partial s_k} F^{(i)}(\mathbf{s}, t)\right) = \sum_{\ell=1}^{K}\Bigg[\frac{\partial^2}{\partial s_j \partial s_k} u^{(\ell)}(\mathbf{s})\frac{\partial}{\partial s_\ell} F^{(i)}(\mathbf{s}, t) + \frac{\partial}{\partial s_j} u^{(\ell)}(\mathbf{s})\frac{\partial^2}{\partial s_\ell \partial s_k} F^{(i)}(\mathbf{s}, t)$$
$$+ \frac{\partial}{\partial s_k} u^{(\ell)}(\mathbf{s})\frac{\partial^2}{\partial s_\ell \partial s_j} F^{(i)}(\mathbf{s}, t) + u^{(\ell)}(\mathbf{s})\frac{\partial^3}{\partial s_\ell \partial s_j \partial s_k} F^{(i)}(\mathbf{s}, t)\Bigg]. \tag{15}$$

We can see from the equation (14) that for any $\ell \in \{1, \cdots, K\}$

$$u^{(\ell)}(\mathbf{1}) = 0$$

$$\frac{\partial}{\partial s_j} u^{(\ell)}(\mathbf{1}) = \begin{cases} \alpha_j - \beta_j & j = \ell \\ \nu_{\ell i} & j \neq \ell \end{cases}$$

$$\frac{\partial^2}{\partial s_j \partial s_k} u^{(\ell)}(\mathbf{1}) = \begin{cases} \nu_{jk} & \ell = j \neq k \\ 2\alpha_j & \ell = j = k. \end{cases}$$

Letting $\mathbf{s} = \mathbf{1}$ and re-arranging equation (15), we can see that

$$\frac{d}{dt}\mathbf{D}_{jk}^{(i)}(t) = \sum_{\ell=1}^{K} \mathbf{A}_{\ell j}\mathbf{D}_{k\ell}^{(i)}(t) + \sum_{\ell=1}^{K} \mathbf{A}_{\ell k}\mathbf{D}_{j\ell}^{(i)}(t) + (1 - \delta_{jk})[\nu_{kj}\mathbf{M}_{ik}(t) + \nu_{jk}\mathbf{M}_{ij}(t)] + 2\delta_{jk}\alpha_j\mathbf{M}_{ij}(t),$$

where $\mathbf{A}$ is the infinitesimal generation matrix, and we know that $\mathbf{D}^{(i)}(t)$ is a symmetric matrix. Thus, we derive the following matrix differential equation

$$\frac{d}{dt}\mathbf{D}^{(i)}(t) = \mathbf{A}^\intercal \mathbf{D}^{(i)}(t) + \mathbf{D}^{(i)}(t)\mathbf{A} + \mathbf{C}^{(i)}(t), \tag{16}$$

where $\mathbf{C}_{jk}^{(i)}(t) = \begin{cases} \nu_{kj}\mathbf{M}_{ik}(t) + \nu_{jk}\mathbf{M}_{ij}(t) & j \neq k \\ 2\alpha_j\mathbf{M}_{ij}(t) & j = k \end{cases}$. Observe that equation (16) is a Lyapunov matrix differential equation. With the initial condition $\mathbf{D}^{(i)}(0) = 0$, we have the solution

$$\mathbf{D}^{(i)}(t) = \exp(t\mathbf{A}^\intercal)\left(\int_0^t \exp(-\tau\mathbf{A}^\intercal)\mathbf{C}^{(i)}(\tau)\exp(-\tau\mathbf{A})d\tau\right)\exp(t\mathbf{A})$$

$$= \int_0^t (\mathbf{M}(t - \tau))^\intercal \mathbf{C}^{(i)}(\tau)(\mathbf{M}(t - \tau))\,d\tau.$$

Then we can derive the covariance $\mathbf{\Xi}^{(i)}(t)$ from $\mathbf{D}^{(i)}(t)$ from equation (13).

6.4 Extension of Proposition 1

In Proposition 1, we assume that time points $\mathcal{T} = \{t_1, \cdots, t_{N_T}\}$ are independent of the initial cell count n. This assumption becomes invalid if we assign specific numerical values to n and $t_1, \cdots t_{N_T}$. Therefore, in this subsection, we aim to relax this assumption to make our results applicable in a more realistic scenario, where there is a dependency between the set of time points and the initial cell count.

To that end, we denote

$$\mathbf{Y}(\mathcal{T}_n) := [Y(t_{1,n}), Y(t_{2,n}), \cdots, Y(t_{N_T,n})],$$

where $\mathcal{T}_n$ depends on n and $t_{i,n}$ belongs to one of the following sets of sequences:

$$\mathcal{F} := \{(t_n)_{n \in \mathbb{N}}; \limsup_{n \to \infty} t_n < \infty\}$$

$$\mathcal{I} := \{(t_n)_{n \in \mathbb{N}}; \lim_{n \to \infty} t_n = \infty\}.$$

The sets $\mathcal{F}$ and $\mathcal{I}$ correspond to the finite time points sequence, which remain finite as $n \to \infty$, and the infinite time points sequence, which diverge to infinite as $n \to \infty$, respectively.

Since we introduce the dependency on n, the covariance between $Z(t_{i,n})$ and $Z(t_{j,n})$ no longer remains constant as n approaches infinity. Therefore, directly applying the multivariate CLT is not valid in this case; instead, we should utilize the more general Lindeberg-Feller multivariate CLT.

To apply the Lindeberg-Feller multivariate CLT, we denote the normalized random vector

$$\bar{\mathbf{Y}}(\mathcal{T}_n) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \bar{\mathbf{y}}_j^{(i)}(\mathcal{T}_n) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} [\exp(-\lambda_1 t_{1,n})\langle \mathbf{B}_j^{(i)}(t_{1,n}), \mathbf{1}\rangle, \cdots, \exp(-\lambda_1 t_{N_T,n})\langle \mathbf{B}_j^{(i)}(t_{N_T,n}), \mathbf{1}\rangle].$$

We next denote the

$$\bar{\boldsymbol{\mu}}^{(i)}(\mathcal{T}_n) = [\exp(-\lambda_1 t_{1,n})\langle \mathbf{m}^{(i)}(t_{1,n}), \mathbf{1}\rangle, \cdots, \exp(-\lambda_1 t_{N_T,n})\langle \mathbf{m}^{(i)}(t_{N_T,n}), \mathbb{1}\rangle]$$

as the expectation of $\bar{\mathbf{y}}_j^{(i)}(\mathcal{T}_n)$. Note that all elements in $\bar{\boldsymbol{\mu}}^{(i)}(\mathcal{T}_n)$ are bounded above by 1 according to the spectral property of the *mean matrix*.

Same as Proposition 1, we define a centered and normalized process for a given $i \in \mathcal{K}$

$$\bar{\mathbf{W}}_{n_i}^{(i)}(\mathcal{T}_n) = \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \left(\bar{\mathbf{y}}_j^{(i)}(\mathcal{T}_n) - \bar{\boldsymbol{\mu}}^{(i)}(\mathcal{T}_n) \right)$$

and claim Proposition 2 from Lindeberg-Feller multivariate CLT

Proposition 2. *as* $n \to \infty$

$$\bar{\mathbf{W}}_{n_i}^{(i)}(\mathcal{T}_n) \Rightarrow N(0, \bar{\mathbf{V}}^{(i)}(\mathcal{T}_n))$$

where for $1 \le a \le b \le N_T$

$$\bar{\mathbf{V}}_{a,b}^{(i)}(\mathcal{T}_n) = \begin{cases} \mathbf{1}\exp(-2\lambda_1 t_{a,n})\mathbf{\Xi}^{(i)}(t_{a,n})\exp(-\lambda_1(t_{b,n} - t_{a,n}))\mathbf{M}(t_{b,n} - t_{a,n})\mathbf{1}^\intercal & t_{a,n}, t_{b,n} \in \mathcal{F} \\ \mathbf{1}\exp(-2\lambda_1 t_{a,n})\mathbf{\Xi}^{(i)}(t_{a,n})\mathbf{P}\mathbf{1}^\intercal & t_{b,n} - t_{a,n} \in \mathcal{I}, t_{a,n} \in \mathcal{F} \\ \mathbf{1}\mathbf{Q}^{(i)}\exp(-\lambda_1(t_{b,n} - t_{a,n}))\mathbf{M}(t_{b,n} - t_{a,n})\mathbf{1}^\intercal & t_{b,n} - t_{a,n} \in \mathcal{F}, t_{a,n} \in \mathcal{I} \\ \mathbf{1}\mathbf{Q}^{(i)}\mathbf{P}\mathbf{1}^\intercal & t_{b,n} - t_{a,n} \in \mathcal{I}, t_{a,n} \in \mathcal{I}. \end{cases}$$

The $\mathbf{Q}^{(i)}$ *and* $\mathbf{P}$ *are the limit of the* $\exp(-2\lambda_1 t)\mathbf{\Xi}^{(i)}(t)$ *and* $\exp(-\lambda_1 t)\mathbf{M}(t)$ *respectively as* $t \to \infty$.

27

Proof. For simplicity, we denote $\mathbf{z}_j^{(i)}(\mathcal{T}_n) = \bar{\mathbf{y}}_j^{(i)}(\mathcal{T}_n) - \bar{\boldsymbol{\mu}}^{(i)}(\mathcal{T}_n)$. It is obvious that $\mathrm{E}\left[\mathbf{z}_j^{(i)}(\mathcal{T}_n)\right] = \mathbf{0}$ for all n. Then we need to check the behavior of its covariance matrix. Without loss of generality, we assume $1 \le a \le b \le N_T$. For a,b-th element of the second moment of $\mathbf{z}_j^{(i)}(\mathcal{T}_n)$:

$$[\mathbb{V}\mathbf{z}_j^{(i)}(\mathcal{T}_n)]_{a,b} = \exp(-\lambda_1(t_{a,n} + t_{b,n}))\mathrm{Cov}\left(\langle \mathbf{B}^{(i)}(t_{a,n}), \mathbf{1}\rangle, \langle \mathbf{B}^{(i)}(t_{b,n}), \mathbf{1}\rangle\right)$$

$$\mathrm{Cov}\left(\langle \mathbf{B}^{(i)}(t_{a,n}), \mathbf{1}\rangle, \langle \mathbf{B}^{(i)}(t_{b,n}), \mathbf{1}\rangle\right) = \mathbf{1}\left\{\mathrm{E}\left[\mathbf{B}^{(i)}(t_{a,n})^\mathsf{T}\mathbf{B}^{(i)}(t_{b,n})\right] - \mathrm{E}\left[\mathbf{B}^{(i)}(t_{a,n})\right]^\mathsf{T}\mathrm{E}\left[\mathbf{B}^{(i)}(t_{b,n})\right]\right\}\mathbf{1}^\mathsf{T}$$

$$= \mathbf{1}\left\{\mathrm{E}\left[\mathbf{B}^{(i)}(t_{a,n})^\mathsf{T}\mathrm{E}\left[\mathbf{B}^{(i)}(t_{b,n}) \mid \mathbf{B}^{(i)}(t_{a,n})\right]\right]\right.$$
$$\left. - \mathrm{E}\left[\mathbf{B}^{(i)}(t_{a,n})\right]^\mathsf{T}\mathrm{E}\left[\mathrm{E}\left[\mathbf{B}^{(i)}(t_{b,n}) \mid \mathbf{B}^{(i)}(t_{a,n})\right]\right]\right\}\mathbf{1}^\mathsf{T}$$

$$= \mathbf{1}\left\{\mathrm{E}\left[\mathbf{B}^{(i)}(t_{a,n})^\mathsf{T}\mathbf{B}^{(i)}(t_{a,n})\mathbf{M}(t_{b,n} - t_{a,n})\right]\right.$$
$$\left. - \mathrm{E}\left[\mathbf{B}^{(i)}(t_{a,n})\right]^\mathsf{T}\mathrm{E}\left[\mathbf{B}^{(i)}(t_{a,n})\mathbf{M}(t_{b,n} - t_{a,n})\right]\right\}\mathbf{1}^\mathsf{T}$$

$$= \mathbf{1}\left\{\mathrm{E}\left[\mathbf{B}^{(i)}(t_{a,n})^\mathsf{T}\mathbf{B}^{(i)}(t_{a,n})\right] - \mathrm{E}\left[\mathbf{B}^{(i)}(t_{a,n})\right]^\mathsf{T}\mathrm{E}\left[\mathbf{B}^{(i)}(t_{a,n})\right]\right\}\mathbf{M}(t_{b,n} - t_{a,n})\mathbf{1}^\mathsf{T}$$

$$= \mathbf{1}\boldsymbol{\Xi}^{(i)}(t_{a,n})\mathbf{M}(t_{b,n} - t_{a,n})\mathbf{1}^\mathsf{T}$$

$$[\mathbb{V}\mathbf{z}_j^{(i)}(\mathcal{T}_n)]_{a,b} = \mathbf{1}\exp(-2\lambda_1 t_{a,n})\boldsymbol{\Xi}^{(i)}(t_{a,n})\exp(-\lambda_1(t_{b,n} - t_{a,n}))\mathbf{M}(t_{b,n} - t_{a,n})\mathbf{1}^\mathsf{T}$$

We know from Athreya and Ney that:

$$\lim_{t \to \infty} \mathbf{M}(t)\exp(-\lambda_1 t) = \mathbf{P}$$

where $\mathbf{P}$ is the outer product of the left and right eigenvector of $\mathbf{M}(t)$ corresponding to eigenvalue $\exp(\lambda_1 t)$, and there exists a finite matrix $\mathbf{Q}^{(i)}$ such that

$$\lim_{t \to \infty} \boldsymbol{\Xi}^{(i)}(t)\exp(-2\lambda_1 t) = \mathbf{Q}^{(i)}.$$

Therefore, we know that all the elements in $\mathbb{V}\mathbf{z}_j^{(i)}(\mathcal{T}_n)$ are finite. In particular, we claim the following condition for Lindeberg-Feller multivariate CLT holds:

$$\lim_{n_i \to \infty} \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbb{V}\mathbf{z}_j^{(i)}(\mathcal{T}_n) = \lim_{n \to \infty} \mathbb{V}\mathbf{z}_j^{(i)}(\mathcal{T}_n) = \bar{\mathbf{V}}^{(i)}(\mathcal{T}_n),$$

where

$$\bar{\mathbf{V}}_{a,b}^{(i)}(\mathcal{T}_n) = \begin{cases} \mathbf{1}\exp(-2\lambda_1 t_{a,n})\boldsymbol{\Xi}^{(i)}(t_{a,n})\exp(-\lambda_1(t_{b,n} - t_{a,n}))\mathbf{M}(t_{b,n} - t_{a,n})\mathbf{1}^\mathsf{T} & t_{a,n}, t_{b,n} \in \mathcal{F} \\ \mathbf{1}\exp(-2\lambda_1 t_{a,n})\boldsymbol{\Xi}^{(i)}(t_{a,n})\mathbf{P}\mathbf{1}^\mathsf{T} & t_{b,n} - t_{a,n} \in \mathcal{I}_\tau \\ \mathbf{1}\mathbf{Q}\exp(-\lambda_1(t_{b,n} - t_{a,n}))\mathbf{M}(t_{b,n} - t_{a,n})\mathbf{1}^\mathsf{T} & t_{a,n} \in \mathcal{I} \\ \mathbf{1}\mathbf{Q}\mathbf{P}\mathbf{1}^\mathsf{T} & t_{b,n} - t_{a,n} \in \mathcal{I}_\tau, t_{a,n} \in \mathcal{I}. \end{cases}$$

The first equality is due to the following two observations: 1. As p_i independent of n, we know that $n_i \to \infty$ is equivalent to $n \to \infty$. 2. For any given n, the random vector $\mathbf{z}_j^{(i)}(\mathcal{T}_n)$ is i.i.d. copy of $\mathbf{z}^{(i)}(\mathcal{T}_n)$.

Then we claim the Lindeberg condition:

$$\lim_{n_i \to \infty} \frac{1}{n_i}\sum_{j=1}^{n_i} \mathrm{E}\left[\|\mathbf{z}_j^{(i)}(\mathcal{T}_n)\|^2 \mathbb{1}(\|\mathbf{z}_j^{(i)}(\mathcal{T}_n)\| \ge \epsilon\sqrt{n})\right] = 0 \tag{17}$$

for a given $\epsilon > 0$, where $\mathbb{1}(\cdot)$ is the indicator function, is satisfied. Due to the aforementioned two observations, it is equivalent to show that

$$\lim_{n\to\infty} \mathrm{E}\left[\|\mathbf{z}^{(i)}(\mathcal{T}_n)\|^2 \mathbb{1}(\|\mathbf{z}^{(i)}(\mathcal{T}_n)\| \geq \epsilon\sqrt{n})\right] = 0. \tag{18}$$

We applied the Cauchy-Schwarz inequality to (18), which simplifies the task to showing that

$$\lim_{n\to\infty} \mathrm{E}\left[\|\mathbf{z}^{(i)}(\mathcal{T}_n)\|^4\right]^{1/2} \mathrm{E}\left[\mathbb{1}(\|\mathbf{z}^{(i)}(\mathcal{T}_n)\| \geq \epsilon\sqrt{n})\right]^{1/2} = 0. \tag{19}$$

Then we argue that

$$\lim_{n\to\infty} \mathrm{E}\left[\|\mathbf{z}^{(i)}(\mathcal{T}_n)\|^4\right] = \lim_{n\to\infty} \mathrm{E}\left[\left[\sum_{\tau=1}^{N_T} \exp(-2\lambda_1 t_{\tau,n})\left(\langle\mathbf{B}^{(i)}(t_{\tau,n}), \mathbf{1}\rangle - \langle\mathbf{m}^{(i)}(t_{\tau,n}), \mathbf{1}\rangle\right)^2\right]^2\right]$$

$$\leq \lim_{n\to\infty} \mathrm{E}\left[\left[\sum_{\tau=1}^{N_T} \exp(-2\lambda_1 t_{\tau,n})(\mathbf{1}\mathbf{B}^{(i)}(t_{\tau,n})^\mathsf{T}\mathbf{B}^{(i)}(t_{\tau,n})\mathbf{1}^\mathsf{T} + \mathbf{1}\mathbf{m}^{(i)}(t_{\tau,n})^\mathsf{T}\mathbf{m}^{(i)}(t_{\tau,n})\mathbf{1}^\mathsf{T})\right]^2\right]$$

$$= \lim_{n\to\infty} \sum_{1\leq\tau\leq\ell\leq N_T} \exp(-2\lambda_1(t_{\tau,n} + t_{\ell,n}))(c_{\tau,\ell}^4 + c_{\tau,\ell}^2 + c_{\tau,\ell}^0),$$

where

$$c_{\tau,\ell}^4 = \mathrm{E}\left[\mathbf{1}\mathbf{B}^{(i)}(t_{\tau,n})^\mathsf{T}\mathbf{B}^{(i)}(t_{\tau,n})\mathbf{1}^\mathsf{T}\mathbf{1}\mathbf{B}^{(i)}(t_{\ell,n})^\mathsf{T}\mathbf{B}^{(i)}(t_{\ell,n})\mathbf{1}^\mathsf{T}\right]$$

$$c_{\tau,\ell}^2 = \mathrm{E}\left[\mathbf{1}\mathbf{B}^{(i)}(t_{\tau,n})^\mathsf{T}\mathbf{B}^{(i)}(t_{\tau,n})\mathbf{1}^\mathsf{T}\right]\mathbf{1}\mathbf{m}^{(i)}(t_{\ell,n})^\mathsf{T}\mathbf{m}^{(i)}(t_{\ell,n})\mathbf{1}^\mathsf{T}$$

$$c_{\tau,\ell}^0 = \mathbf{1}\mathbf{m}^{(i)}(t_{\tau,n})^\mathsf{T}\mathbf{m}^{(i)}(t_{\tau,n})\mathbf{1}^\mathsf{T}\mathbf{1}\mathbf{m}^{(i)}(t_{\ell,n})^\mathsf{T}\mathbf{m}^{(i)}(t_{\ell,n})\mathbf{1}^\mathsf{T}.$$

With Lemma 1 and the fact $\lim_{t\to\infty} \exp(-2\lambda_1 t)\mathbf{1}\mathbf{m}^{(i)}(t)^\mathsf{T}\mathbf{m}^{(i)}(t)\mathbf{1}^\mathsf{T} < \infty$, we can directly see that

$$\lim_{n\to\infty} \exp(-2\lambda_1(t_{\tau,n} + t_{\ell,n}))c_{\tau,\ell}^0 < \infty$$

$$\lim_{n\to\infty} \exp(-2\lambda_1(t_{\tau,n} + t_{\ell,n}))c_{\tau,\ell}^2 < \infty.$$

For $c_{\tau,\ell}^4$, we apply the Cauchy-Schwarz inequality to obtain

$$\exp(-2\lambda_1(t_{\tau,n} + t_{\ell,n}))c_{\tau,\ell}^4 \leq \left[\exp(-4\lambda_1 t_{\tau,n})c_{\tau,n}^4\right]^{1/2}\left[\exp(-4\lambda_1 t_{\ell,n})c_{\ell,n}^4\right]^{1/2},$$

where

$$c_{\tau,n}^4 = \mathrm{E}\left[(\mathbf{1}\mathbf{B}^{(i)}(t_{\tau,n})^\mathsf{T}\mathbf{B}^{(i)}(t_{\tau,n})\mathbf{1}^\mathsf{T})^2\right], \text{ for } 1 \leq \tau \leq N_T.$$

From Lemma 1, we claim that for $1 \leq \tau \leq N_T$

$$\lim_{n\to\infty} \exp(-4\lambda_1 t_{\tau,n})c_{\tau,n}^4 < \infty.$$

Consequently, we have

$$\lim_{n\to\infty} \mathrm{E}\left[\|\mathbf{z}^{(i)}(\mathcal{T}_n)\|^4\right] < \infty.$$

29

Next, we observe that

$$\mathrm{E}\Big[\mathbb{1}(\|\mathbf{z}^{(i)}(\mathcal{T}_n)\| \geq \epsilon\sqrt{n})\Big] = \mathbb{P}(\|\mathbf{z}^{(i)}(\mathcal{T}_n)\| \geq \epsilon\sqrt{n}) \leq \frac{\mathrm{E}\big[\|\mathbf{z}^{(i)}(\mathcal{T}_n)\|\big]}{\epsilon\sqrt{n}}.$$

As we have already shown that $\lim_{n\to\infty} \mathrm{E}\big[\|\mathbf{z}^{(i)}(\mathcal{T}_n)\|^4\big] < \infty$, we can directly show that

$$\lim_{n\to\infty} \frac{\mathrm{E}\big[\|\mathbf{z}^{(i)}(\mathcal{T}_n)\|\big]}{\epsilon\sqrt{n}} = 0.$$

As a result, we have equation (19) follows.

We conclude this proof by applying the Lindeberg multivariate CLT:

$$\mathbf{W}_{n_i}^{(i)}(\mathcal{T}_n) = \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \mathbf{z}_j^{(i)}(\mathcal{T}_n) \Rightarrow N(0, \bar{\mathbf{V}}^{(i)}(\mathcal{T}_n)).$$

$\square$

In Proposition 2, we demonstrate with an appropriate normalization factor the normalized and centered random vector $\mathbf{Y}$ is still approximately normally distributed when time depends on n. In practice, because we cannot employ $n \to \infty$, Proposition 1 is enough to obtain a good approximation of $\mathbf{Y}$.

6.5 Details of data generation and optimization implementation in the *in silico* experiments

To describe the data generation and optimization in the *in silico* experiment, we define two distinct ranges:

1. the *Generating parameters range* based on the assumption outlined in Section 2.5.

2. the *Optimization feasible range* employed in all of the *in silico* experiments.

These two ranges characterize most of the differences among the *in silico* experiments, while the procedure of the experiments is the same. The detailed *in silico* experiments procedure is given by the following. We first randomly select a 'true parameters set' from the *Generating parameters range* to simulate the *in silico* data through the Gillespie algorithm [5]. Subsequently, we solve the MLE problem in equation (7) within the *Optimization feasible range*. In particular, we employ the MATLAB Optimization Toolbox [21] function *fmincon* with sequential quadratic programming (sqp) solver. Given the complex likelihood equation, we solve the optimization starting from 20 initial points within the feasible region and select the estimation with the lowest negative log-likelihood.

We denote the 'true parameters sets' and the estimated parameters as $\boldsymbol{\theta}^*$ and $\hat{\boldsymbol{\theta}}$ respectively, while referring to the *Generating parameters range* and *Optimization feasible range* as $\boldsymbol{\Theta}^*$ and $\hat{\boldsymbol{\Theta}}$. These ranges simulate realistic scenarios and reflect our prior knowledge about nature. One principle in setting these two ranges is that $\boldsymbol{\Theta}^* \subseteq \hat{\boldsymbol{\Theta}}$ to simulate insufficient prior knowledge about the 'true nature'.

6.5.1   Base experiment

In this experiment, we adhere to the assumptions outlined in Section 2.5. In particular, the *Generating parameter range* $\boldsymbol{\Theta}^*$ and the *Optimization feasible range* satisfy the assumptions established in Section 2.5. For simplicity, we implemented a special case of Assumption 1: $\alpha_s \in (\beta_s, \beta_s + 0.1)$ and $\alpha_d = \beta_d$, (i.e. $0 = \kappa_d \le \kappa_s \in (0, 0.1)$). Nevertheless, we also tested for the case when $\kappa_d > 0$ and got a similar results. Assumption 2 can be directly implemented in the selection of $\boldsymbol{\Theta}^*$ and $\hat{\boldsymbol{\Theta}}$, and Assumption 5 can be implemented in the computation of likelihood. In accordance with Assumption 4, where $b_{s,\beta} = b_{s,\nu} = 1$, the selection of $E_{s,\beta}$ and $E_{s,\nu}$ does not impact the likelihood computation, so we set $E_{s,\beta} = E_{s,\nu} = 1$ for simplicity. Conversely, we carefully choose the values for $E_{d,\beta}$ and $E_{d,\nu}$ within the concentration levels implemented in the *in silico* experiment to ensure the drug has an observable effect on the tumor within the given concentration levels.

We conclude our selection of $\boldsymbol{\Theta}^*$ and $\hat{\boldsymbol{\Theta}}$ in Table 8. To prevent the occurrence of unreasonably large net growth rates during the optimization process, we imposed two linear constraints: $\alpha_s - \beta_s \le 0.1$ and $\nu_{sd} + \alpha_d - \beta_d \le 0.5$. In addition to the subpopulation parameters, we also selected the standard deviation parameter $c \in (0, 10)$ for the observation noised added to the data.

Parameters	$\boldsymbol{\Theta}^*$	$\hat{\boldsymbol{\Theta}}$
α_s	$(\beta_s, \beta_s + 0.1)$	$(0, 1)$
β_s	$(10^{-3}, 0.9)$	$(0, 1)$
ν_{sd}	$(0, 0.5)$	$(0, 1)$
α_d	β_d	$(0, 1)$
β_d	$(10^{-3}, 0.5)$	$(0, 1)$
$b_{d,\beta}$	$(0.8, 0.9)$	$(0.5, 1)$
$E_{d,\beta}$	$(0.0625, 2.5)$	$(0, 3)$
$b_{d,\nu}$	$(1, 1.1)$	$(1, 1.5)$
$E_{d,\nu}$	$(0.0625, 2.5)$	$(0, 3)$

Table 8: *Generating parameter range* $\boldsymbol{\Theta}^*$ and *Optimization feasible range* $\hat{\boldsymbol{\Theta}}$. According to assumptions in Section 2.5, $b_{s,\beta} = E_{s,\beta} = b_{s,\nu} = E_{s,\nu} = 1$ and $\nu_{d,s} = 0$ are set to both range.

6.5.2   Relaxed drug effect assumption

To model a scenario where the drug has less significant effects on the CSCs, we selected $1 - b_{s,\beta} < 1 - b_{d,\beta}$ and $b_{s,\nu} - 1 < b_{d,\nu} - 1$. As in the base experiment, we present the new *Generating parameters range* and *Optimization feasible range* in Table 9.

6.5.3   Relaxed initial proportion assumption

Compared to the base experiment, we added initial proportion as variables to be estimated in these experiments. The corresponding *Generating parameters range* and *Optimization feasible range* are given in Table 10. It is important to note that during the optimization process, we also apply an equality constraint, $p_d = 1 - p_s$, to avoid unrealistic initial proportion.

6.6   Bootstrapping technique

To compute the confidence interval (CIs) of estimation, we employ the bootstrapping technique. This technique requires resampling the original dataset to obtain a new dataset. In Section 3.1.1,

Parameters	Θ^*	$\hat{\Theta}$
α_s	$(\beta_s, \beta_s + 0.1)$	$(0, 1)$
β_s	$(10^{-3}, 0.9)$	$(0, 1)$
ν_{sd}	$(0, 0.5)$	$(0, 1)$
$b_{s,\beta}$	$(0.97, 1)$	$(0.5, 1)$
$E_{s,\beta}$	$(0.0625, 2.5)$	$(0, 5)$
$b_{s,\nu}$	$(1, 1.03)$	$(1, 1.5)$
$E_{s,\nu}$	$(0.0625, 2.5)$	$(0, 5)$
α_d	β_d	$(0, 1)$
β_d	$(10^{-3}, 0.5)$	$(0, 1)$
$b_{d,\beta}$	$(0.8, 0.9)$	$(0.5, 1)$
$E_{d,\beta}$	$(0.0625, 2.5)$	$(0, 3)$
$b_{d,\nu}$	$(1, 1.1)$	$(1, 1.5)$
$E_{d,\nu}$	$(0.0625, 2.5)$	$(0, 3)$

Table 9: *Generating parameter range* Θ^* and *Optimization feasible range* $\hat{\Theta}$ for relaxed drug effect experiments. According to assumptions in Section 2.5, $\nu_{d,s} = 0$ is set to both ranges.

we resampled 13 replicates from 20 generated replicates, while in Appendix 6.10 we resampled 10 replicates from 14 replicates. Then the MLE process was conducted to obtain a bootstrap estimate. Based on 100 bootstrap estimates, we obtained the 95% CIs by taking the 2.5 and 97.5 percentile of those estimates.

6.7 Parameters of interest

To evaluate the quality of the estimations, we consider the estimations of the following quantity, which can profile the phenotypical switching dynamics and the subpopulation sensitivity to the drug.

- **Stable proportion**:

  The *stable proportion* of the multi-type branching process is determined by the normalized left eigenvector, $\boldsymbol{\pi}$, corresponding to the largest eigenvalue, $e^{\lambda_1 t}$, of the *mean matrix*. As a result, it depends on the symmetric birth rates, death rates, and asymmetric birth rates of each subpopulation. In our drug-effect model, we assume that either the death rates or the asymmetric birth rates of each subpopulation are affected by the drug, and thus, the *stable proportion* also depends on the drug concentration levels. To better understand the nature of the given tumor, we consider the *stable proportion* as one important parameter of interest.

- GR_{50}:

  The GR_{50}, introduced in [9], is a newly proposed summary metric of drug-sensitivity. It is defined as the concentration at which a drug's effect on cell growth is half the observed effect. For a given maximum concentration level, d_m, applied during the experiment and corresponding drug effect parameters, (b, E), we compute the GR_{50} by

  $$GR_{50} = E\left(\frac{1 - e^{r_m}}{e^{r_m} - b}\right),$$

  where $r_m = \frac{1}{2}\log\left(b + \frac{1-b}{1 + \frac{d_m}{E}}\right)$ is the half-maximum effects.

Parameters	Θ^*	$\hat{\Theta}$
p_s	$(0, 1)$	$(0, 1)$
α_s	$\beta_s, \beta_s + 0.1$	$(0, 1)$
β_s	$(10^{-3}, 0.9)$	$(0, 1)$
ν_{sd}	$(0, 0.5)$	$(0, 1)$
p_d	$1 - p_s$	$(0, 1)$
α_d	β_d	$(0, 1)$
β_d	$(10^{-3}, 0.5)$	$(0, 1)$
$b_{d,\beta}$	$(0.8, 0.9)$	$(0.5, 1)$
$E_{d,\beta}$	$(0.0625, 2.5)$	$(0, 3)$
$b_{d,\nu}$	$(1, 1.1)$	$(1, 1.5)$
$E_{d,\nu}$	$(0.0625, 2.5)$	$(0, 3)$

Table 10: *Generating parameter range* Θ^* and *Optimization feasible range* $\hat{\Theta}$ for relaxed initial proportion experiments. According to assumptions in Section 2.5, $b_{s,\beta} = E_{s,\beta} = b_{s,\nu} = E_{s,\nu} = 1$ and $\nu_{d,s} = 0$ are set to both range.

6.8 Detailed analysis of limited division CNSCs dynamics assumption

To better understand how the our framework analyzes the data with limited CNSCs symmetric division, we monitor the relative difference (RD) between the estimation and true parameters, defined as

$$Er(\hat{x}; x^*) = \frac{\hat{x} - x^*}{x^*}. \tag{20}$$

Compared to RE, RD can better depict overestimation and underestimation. As an analogy to the $RE = 0.2$ threshold, we consider $RD = 0.2$ and $RD = -0.2$ lines as thresholds for accurate estimation. The RD of the estimation from 30 experiments with limited division number $G = 0, 1, 2, 3$ are shown in Figure 11, Figure 12, Figure 13, and Figure 14, respectively.

When $G = 0$, our novel framework accurately captures the limited division cell dynamics by correctly identifying the CNSCs' symmetric division rate $\alpha_d = 0$. Therefore, in Figure 11, we observe that every parameter ehibits an RD centered around 0. However, there are many instances where the estimations of other parameters related to CNSCs have RD value less than -0.2 or greater than 0.2. We suggest that this could be due to the small population of CNSCs compared to CSCs when $G = 0$; hence, the estimation quality of these parameters might be influenced by the variability across distinct experiments.

In Figures 12 to 14, we observe that the estimation of α_d becomes more accurate with increasing G. However, the differentiation rate ν_{sd} and CNSCs' death rate β_d are often overestimated. Conversely, the CSCs' symmetric division rate α_s and death rate β_s tend to be slightly underestimated. Overall, we conclude that our framework systematically overestimates ν_{sd} and β_d, while underestimating α_s, β_s in order to better capture the α_d.

Surprisingly, the estimation of $E_{d,\beta}$ and $E_{d,\nu}$ are not systematically overestimated or underestimated. Additionally, these two quantities becomes more accurately estimated with increasing G. Therefore, we propose two hypotheses about the estimations of $E_{d,\beta}$ and $E_{d,\nu}$:

1. These estimations are not significantly affected by changes in the underlying tumor growth dynamics.

2. The quality of these estimations might be related to the size of the targeted subpopulation, CNSCs.
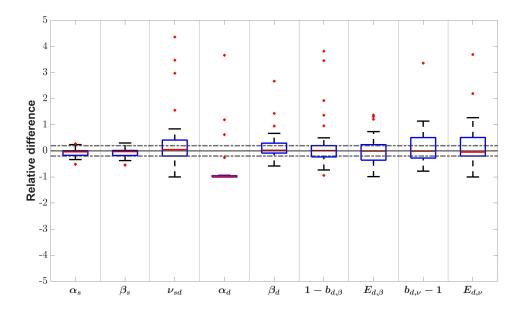
Figure 11: Relative difference of parameter estimation when CNSCs can only symmetrically divide $G = 0$ times. The solid line indicates $RD = 0$ and the dashed line indicates $RD = 0.2$ and $RD = -0.2$.

These conjectures suggest that the drug effects' inflection points $E_{d,\beta}$ and $E_{d,\nu}$ could exhibit robust estimation characteristics across different scenarios, particularly when the targeted subpopulation is abundant.

6.9   Detailed implementation of *in vitro* experiment

In this subsection, we present a detailed implementation of the *in vitro* experiment described in Section 4.

6.9.1   Analysis of the *AGS Time* experiment data

To understand the *AGS Time* experiment data obtained from [25], we visualize the total cell count data at different time points in Figure 15. It is evident that, before time point 12, the total cell count growth dynamics remains indifferent across three varying drug concentration levels. However, after 24 hours, the effect of the drug becomes significant. Therefore, we conclude that there is a time delay before CPX-O can fully activate to affect the AGS cell line.

In contrast to practical observation, our statistical framework does not assume heterogeneity of the drug effect over time. As a result, our current framework struggles to adequately capture the *AGS Time* experiment data without additional assumptions to incorporate the time effect on drug activation. We then consider modeling the time effect on drug activation as one of the future directions, which can further strengthen the current statistical framework. Nevertheless, we are still incorporating some aspects of the information obtained from the *AGS Time* experiment.
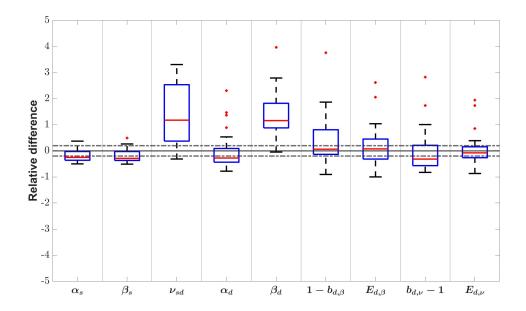
Figure 12: Relative difference of parameter estimation when CNSCs can only symmetrically divide $G = 1$ times. The solid line indicates $RD = 0$ and the dashed line indicates $RD = 0.2$ and $RD = -0.2$.

6.9.2 *AGS Conc* TC data imputation for varying concentration levels at 12 and 24 hours.

Due to the scarcity of the *AGS Time* data, we proposed to employ the simpler exponential growth model to characterize the cell growth dynamics to estimate the TC of *AGS Conc* for varying concentration levels at 12 and 24 hours. In particular, this model comprises two parameters λ, X_0, representing the growth rate and initial cell count, respectively. One can mathematically formulate the cell count $X(t)$ at time t by

$$X(t, d) = X_0 \exp(\lambda(d)t), \tag{21}$$

where the growth rate λ may depend on the drug concentration level d.

As mentioned in the main text, the *AGS Time* data comprises only 3 concentration levels, which is insufficient for estimating 4 drug-related parameters $b_{d,\beta}, E_{d,\beta}, b_{d,\nu}, E_{d,\nu}$. Therefore, we need to utilize the *AGS Conc* data, which collects data at 9 varying concentration levels.

Time-delayed drug effect assumption:

In Appendix 6.9.1, we observed that the cell growth dynamics did not change at the first 12 hours under varying doses of CPX-O. To validate our statistical framework, which does not assume a delay in the drug effect over time, we propose focusing on the data after the 12-hour checkpoint. In other words, we assume that the CPX-O does not have an effect in the first 12 hours.

Shared underlying dynamics between *AGS Time* and *AGS Conc* data assumption:

Because the *AGS Time* and *AGS Conc* data have different initial experimental settings, such as varying the initial total cell count and the proportion between CSCs and CNSCs, we need to assume that under no drug effect (DMSO), the underlying growth dynamics are consistent for these
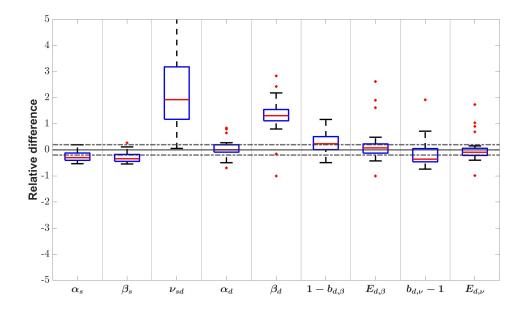
Figure 13: Relative difference of parameter estimation when CNSCs can only symmetrically divide $G = 2$ times. The solid line indicates $RD = 0$ and the dashed line indicates $RD = 0.2$ and $RD = -0.2$.

two experiments. Based on this assumption, we can estimate the initial TC ($X_{Conc}(0,0)$) in the *AGS Conc* data from TC at 48 hours ($X_{Conc}(48,0)$) based on the ratio between the initial TC ($X_{Time}(0,0)$) and TC at 48 hours ($X_{Time}(48,0)$) in *AGS Time* data. In other words, we assume the following equation holds

$$\frac{X_{Time}(48,0)}{X_{Time}(0,0)} = \frac{X_{Conc}(48,0)}{X_{Conc}(0,0)}, \tag{22}$$

where the colored quantity is estimated data.

Estimating the *AGS Conc* data at 12 hour

We estimate the $X_{Time}(0,0)$ data by fitting the exponential growth model to the DMSO *AGS Time* data at $2, 6$, and 12 time-points. After obtaining the estimation for $X_{Conc}(0,0)$, we further estimate the $X_{Conc}(12,0)$ based on the exponential growth model fitted to the DMSO *AGS Time* data at $2, 6, 12$ time-points. Since we assume that CPX-O has no effect in the first 12 hours, we assign the same value of $X_{Conc}(12,d) = X_{Conc}(12,0)$ to all nine varying concentration levels. It is also worth mentioning that we estimate $X_{Conc}(12,0)$ separately for each replicate.

Estimating the *AGS Conc* data at 24 hours

To estimate the *AGS Conc* data at 24 hours, we first note that CPX-O affected cell growth dynamics during the time interval $(12, 24$ hours$)$. Therefore, we need to estimate the drug effect during the time interval $(12, 24$ hours$)$. Since we only have three concentration levels, we proposed to utilize the Hill equation with two cytotoxic effects parameters b, E to model the drug effect here. It is also worth noticing that the cytotoxic effect is the main effect that is observed. Based on the exponential growth model, we obtain 3 estimated growth rates $\lambda(0), \lambda(4), \lambda(8)$ for 3 concentration
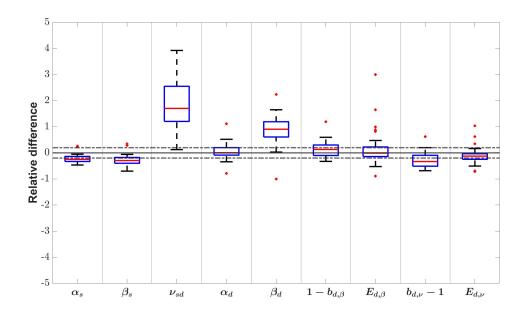
Figure 14: Relative difference of parameter estimation when CNSCs can only symmetrically divide $G = 3$ times. The solid line indicates $RD = 0$ and the dashed line indicates $RD = 0.2$ and $RD = -0.2$.
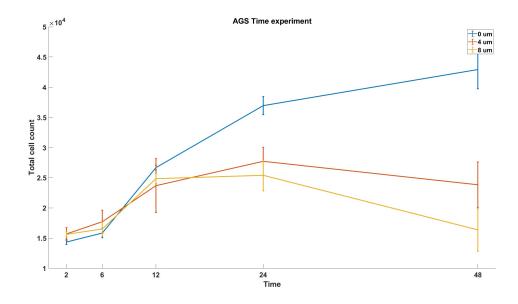

Figure 15: *AGS Time* experiment data.

37

levels of CPX-O. We then fit these growth rates through the 2 parameters Hill equations

$$\lambda(d) = \lambda(0) + \log\left(b + \frac{1-b}{1+d/E}\right).$$

With the estimated parameters b, E, we estimate the growth rate $\lambda(d), d \in \{0, 0.125, 0.25, 0.5, 1, 2, 4, 8, 16\}$. We then plug in these estimated growth rates to estimate the TC at 24 hours, i.e.

$$X(24, d) = X(12, 0)\exp(12\lambda(d)), d \in \{0, 0.125, 0.25, 0.5, 1, 2, 4, 8, 16\}.$$

6.9.3 Details of maximum likelihood estimation

To calculate the AIC value, we first need to obtain the maximum likelihood estimate for each model. Following the experimental setup described in Appendix 6.5, Table 11 provides the *Optimization feasible range* used in the experiment. The same optimization range is applied to the free parameters across all four model assumptions.

Parameters	$\hat{\Theta}$
α_s	$(\beta_s, \beta_s + 0.1)$
β_s	$(0, 1)$
ν_{sd}	$(0, 1)$
α_d	$(\beta_d, \beta_d + 0.1)$
β_d	$(0, 1)$
$b_{d,\beta}$	$(0, 1)$
$E_{d,\beta}$	$(0, 16)$
$m_{d,\beta}$	$(0.1, 5)$
$b_{d,\nu}$	$(1, 2)$
$E_{d,\nu}$	$(0, 16)$
$m_{d,\nu}$	$(0.1, 5)$
k	$(0, 1)$
t_0	$(0, 36)$
c	$(0, 300)$

Table 11: *Optimization feasible range* $\hat{\Theta}$ for *in vitro* drug-induced plasticity experiment.

6.10 Inferring no drug-induced phenotypic transition in imatinib-sensitive and -resistant *Ba/F3* cells data

We further validated our framework using another HTS *in vitro* dataset, which includes a mixture of imatinib-sensitive and -resistant *Ba/F3* cells [16]. This dataset includes the total cell count of the mixture under 11 different concentration levels of imatinib. Each concentration level was evaluated with 14 sample replicates, and cell counts were recorded at 14 time points using live-cell imaging. The specific concentration levels and time points are as follows:

$$\mathcal{D}_{Ba/F3} = \{0, 0.0313, 0.0625, 0.125, 0.25, 0.375, 0.5, 1.25, 2.5, 3.75, 5\}(\mu m),$$

$$\mathcal{T}_{Ba/F3} = \{0, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36\}(\text{hours}).$$

38

In this study, we focused on a dataset named BF21, which corresponds to an initial proportion $2:1$ between the sensitive and resistant subpopulations.

In this case, the drug affects both the resistant and sensitive subpopulations. We therefore adopted model assumptions similar to those in Section 3.2.1. However, unlike the model in Section 3.2.1, we allow ν_{sr} to be nonzero, indicating potential natural plasticity of the sensitive cells. As previously noted, the estimation quality was suboptimal in Section 3.2.1. To improve this, we fixed the initial proportion between the sensitive and resistant subpopulations at $2:1$. The model used in this experiment includes 19 free parameters:

$$\theta = (\alpha_s, \beta_s, \nu_{sr}, b_{s,\beta}, E_{s,\beta}, m_{s,\beta}, b_{s,\nu}, E_{s,\nu}, m_{s,\nu}, \alpha_r, \beta_r, \nu_{rs}, b_{r,\beta}, E_{r,\beta}, m_{r,\beta}, b_{r,\nu}, E_{r,\nu}, m_{r,\nu}, c),$$

where the subscript s and r represent the sensitive and resistant subpopulations, respectively. A detailed implementation of the *Optimization feasible region*, $\hat{\Theta}$, is provided in Table 12.

Parameters	$\hat{\Theta}$
α_s	$(\beta_s, \beta_s + 0.06)$
β_s	$(0, 1)$
ν_{sr}	$(0, 0.1)$
$b_{s,\beta}$	$(0.878, 1)$
$E_{s,\beta}$	$(0, 50)$
$m_{s,\beta}$	$(0.001, 20)$
$b_{s,\nu}$	$(1, 2)$
$E_{s,\nu}$	$(0, 50)$
$m_{s,\nu}$	$(0.001, 20)$
α_r	$(\beta_r, \beta_r + 0.06)$
β_r	$(0, 1)$
ν_{rs}	$(0, 0.1)$
$b_{r,\beta}$	$(0.878, 1)$
$E_{r,\beta}$	$(0, 50)$
$m_{r,\beta}$	$(0.001, 20)$
$b_{r,\nu}$	$(1, 2)$
$E_{r,\nu}$	$(0, 50)$
$m_{r,\nu}$	$(0.001, 20)$
c	$(0, 100)$

Table 12: *Optimization feasible range* $\hat{\Theta}$ for relaxed drug effect experiments. The subscript s, r represent parameters for sensitive and resistant subpopulations respectively.

In our previous work [35], we performed parameter estimations on this dataset using a simpler version of the current framework that did not account for potential subpopulation transitions. This simpler version aligned with our prior belief that these transitions were very rare. Here, we aimed to determine whether our more complex framework could yield similar conclusions. For assessing transition effects, we analyzed the estimated asymmetrical division rates ν_{sr} and ν_{rs}. Since the model can only identify observed drug effects, we use the drug effects on the asymmetrical division rate at the maximum concentration level ($d = 5$) to evaluate drug-induced transition effects, denoted as $DIT_{sr} := \nu_{sr}(5) - \nu_{sr}$ and $DIT_{rs} := \nu_{rs}(5) - \nu_{rs}$.

Due to potential numerical error and inexact optimization solutions, we report in Table 13 the number of bootstrap estimates below 10^{-6} for $(\nu_{sr}, \nu_{rs}, DIT_{sr}, DIT_{rs})$ out of 100 bootstrap

estimates. The estimates of the drug-induced transition effects (DIT_{sr}, DIT_{rs}) are all below 10^{-6}, and most of estimates of ν_{rs} are also below 10^{-6}. These results suggest that drug-induced transitions between cell types are very unlikely. However, there are only 6 bootstrap instances where ν_{sr} is below 10^{-6}. We hypothesize that this discrepancy is due to either a genuine natural transition from the sensitive to the resistant subpopulation, or an artifact of noise in the initial proportions that is not adequately captured by our model, which assumes a fixed initial proportion. In either case, our framework indicates that drug-induced transitions do not play a large role in the dynamics.

ν_{sr}	ν_{rs}	DIT_{sr}	DIT_{rs}
6	80	100	100

Table 13: Number of bootstrap estimates below 10^{-6} for $(\nu_{sr}, \nu_{rs}, DIT_{sr}, DIT_{rs})$ among 100 bootstrap estimates.

References

[1] Krishna B Athreya and Peter E Ney. *Branching processes*. Courier Corporation, 2004.

[2] Soufiane Boumahdi and Frederic J de Sauvage. The great escape: tumour cell plasticity in resistance to targeted therapy. *Nature reviews Drug discovery*, 19(1):39–56, 2020.

[3] David Dingli and Franziska Michor. Successful therapy must eradicate cancer stem cells. *Stem cells*, 24(12):2603–2610, 2006.

[4] Anusha Dravid, Brad Raos, Darren Svirskis, and Simon J O'Carroll. Optimised techniques for high-throughput screening of differentiated sh-sy5y cells and application for neurite outgrowth assays. *Scientific reports*, 11(1):23935, 2021.

[5] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.

[6] Aaron Goldman, Biswanath Majumder, Andrew Dhawan, Sudharshan Ravi, David Goldman, Mohammad Kohandel, Pradip K Majumder, and Shiladitya Sengupta. Temporally sequenced anticancer drugs overcome adaptive resistance by targeting a vulnerable chemotherapy-induced phenotypic transition. *Nature communications*, 6(1):6139, 2015.

[7] Einar Bjarki Gunnarsson, Subhajyoti De, Kevin Leder, and Jasmine Foo. Understanding the role of phenotypic switching in cancer drug resistance. *Journal of theoretical biology*, 490:110162, 2020.

[8] Einar Bjarki Gunnarsson, Jasmine Foo, and Kevin Leder. Statistical inference of the rates of cell proliferation and phenotypic switching in cancer. *Journal of Theoretical Biology*, 568:111497, 2023.

[9] Marc Hafner, Mario Niepel, Mirra Chung, and Peter K Sorger. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nature methods*, 13(6):521–527, 2016.

[10] Leonard A Harris, Peter L Frick, Shawn P Garbett, Keisha N Hardeman, B Bishal Paudel, Carlos F Lopez, Vito Quaranta, and Darren R Tyson. An unbiased metric of antiproliferative drug effect in vitro. *Nature methods*, 13(6):497–500, 2016.

[11] Masahiro Hitomi, Anastasia P Chumakova, Daniel J Silver, Arnon M Knudsen, W Dean Pontius, Stephanie Murphy, Neha Anand, Bjarne W Kristensen, and Justin D Lathia. Asymmetric cell division promotes therapeutic resistance in glioblastoma stem cells. *JCI insight*, 6(3), 2021.

[12] Francisco J Iborra, Cristina Martí, Virtu Calabuig-Navarro, Petros Papadopoulos, Salvador Meseguer, Pedro M Iborra, Francisco García, Antonio Martínez-Lorente, Fernando Almazán, and Juana Calabuig. Chemotherapy induces cell plasticity; controlling plasticity increases therapeutic response. *Signal Transduction and Targeted Therapy*, 8(1):256, 2023.

[13] Banu Iskender, Kenan Izgi, and Halit Canatan. Reprogramming bladder cancer cells for studying cancer initiation and progression. *Tumor Biology*, 37:13237–13245, 2016.

[14] Alexandra Jilkine. Mathematical models of stem cell differentiation and dedifferentiation. *Current Stem Cell Reports*, 5:66–72, 2019.

[15] Yohan Kim, Jaemin Jeong, and Dongho Choi. Small-molecule-mediated reprogramming: a silver lining for regenerative medicine. *Experimental & Molecular Medicine*, 52(2):213–226, 2020.

[16] Alvaro Köhn-Luque, Even Moa Myklebust, Dagim Shiferaw Tadele, Mariaserena Giliberto, Leonard Schmiester, Jasmine Noory, Elise Harivel, Polina Arsenteva, Shannon M Mumenthaler, Fredrik Schjesvold, et al. Phenotypic deconvolution in heterogeneous cancer cell populations using drug-screening data. *Cell Reports Methods*, 3(3), 2023.

[17] L Kyjacova, S Hubackova, K Krejcikova, R Strauss, H Hanzlikova, R Dzijak, T Imrichova, J Simova, M Reinis, J Bartek, et al. Radiotherapy-induced plasticity of prostate cancer mobilizes stem-like non-adherent, erk signaling-dependent cells. *Cell Death & Differentiation*, 22(6):898–911, 2015.

[18] Kevin Leder, Ken Pitter, Quincey LaPlant, Dolores Hambardzumyan, Brian D Ross, Timothy A Chan, Eric C Holland, and Franziska Michor. Mathematical modeling of pdgf-driven glioblastoma reveals optimized radiation dosing schedules. *Cell*, 156(3):603–616, 2014.

[19] Yuan Li, Zhenning Wang, Jaffer A Ajani, and Shumei Song. Drug resistance and cancer stem cells. *Cell Communication and Signaling*, 19:1–11, 2021.

[20] Deepti Mathur, Ethan Barnett, Howard I Scher, and Joao B Xavier. Optimizing the future: how mathematical models inform treatment schedules for cancer. *Trends in cancer*, 8(6):506–516, 2022.

[21] Matlab optimization toolbox, R2022a. The MathWorks, Natick, MA, USA.

[22] Sean J Morrison and Judith Kimble. Asymmetric and symmetric stem-cell divisions in development and cancer. *nature*, 441(7097):1068–1074, 2006.

[23] Michael D Nicholson, David Cheek, and Tibor Antal. Sequential mutations in exponentially growing populations. *PLOS Computational Biology*, 19(7):e1011289, 2023.

[24] Thomas A Nketia, Heba Sailem, Gustavo Rohde, Raghu Machiraju, and Jens Rittscher. Analysis of live cell images: Methods, tools and opportunities. *Methods*, 115:65–79, 2017.

[25] Diana Pádua, Paula Figueira, Mariana Pinto, André Filipe Maia, Joana Peixoto, Raquel T Lima, António Pombinho, Carlos Filipe Pereira, Raquel Almeida, and Patrícia Mesquita. High-throughput drug screening revealed that ciclopirox olamine can engender gastric cancer stem-like cells. *Cancers*, 15(17):4406, 2023.

[26] Angela Oliveira Pisco, Amy Brock, Joseph Zhou, Andreas Moor, Mitra Mojtahedi, Dean Jackson, and Sui Huang. Non-darwinian dynamics in therapy-induced cancer drug resistance. *Nature communications*, 4(1):2467, 2013.

[27] Heino Prinz. Hill coefficients, dose–response curves and allosteric mechanisms. *Journal of chemical biology*, 3:37–44, 2010.

[28] Hamed Rezayatmand, Mahboobeh Razmkhah, and Iman Razeghian-Jahromi. Drug resistance in cancer therapy: the pandora's box of cancer stem cells. *Stem cell research & therapy*, 13(1):181, 2022.

[29] Mariangela Russo, Simone Pompei, Alberto Sogari, Mattia Corigliano, Giovanni Crisafulli, Alberto Puliafito, Simona Lamba, Jessica Erriquez, Andrea Bertotti, Marco Gherardi, et al. A modified fluctuation-test framework characterizes the population dynamics and mutation rate of colorectal cancer persister cells. *Nature Genetics*, 54(7):976–984, 2022.

[30] Shigeo Saito, Ying-Chu Lin, Yukio Nakamura, Richard Eckner, Kenly Wuputra, Kung-Kai Kuo, Chang-Shen Lin, and Kazunari K Yokoyama. Potential application of cell reprogramming techniques for cancer research. *Cellular and molecular life sciences*, 76:45–65, 2019.

[31] Zhen-Duo Shi, Kun Pang, Zhuo-Xun Wu, Yang Dong, Lin Hao, Jia-Xin Qin, Wei Wang, Zhe-Sheng Chen, and Cong-Hui Han. Tumor cell plasticity in targeted therapy-induced resistance: mechanisms and new strategies. *Signal Transduction and Targeted Therapy*, 8(1):113, 2023.

[32] Yang Song, Jennifer Soto, Pingping Wang, Qin An, Xuexiang Zhang, SoonGweon Hong, Luke P Lee, Guoping Fan, Li Yang, and Song Li. Asymmetric cell division of fibroblasts is an early deterministic step to generate elite cells during cell reprogramming. *Advanced Science*, 8(7):2003516, 2021.

[33] Yapeng Su, Wei Wei, Lidia Robert, Min Xue, Jennifer Tsoi, Angel Garcia-Diaz, Blanca Homet Moreno, Jungwoo Kim, Rachel H Ng, Jihoon W Lee, et al. Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. *Proceedings of the National Academy of Sciences*, 114(52):13679–13684, 2017.

[34] Kazutoshi Takahashi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, and Shinya Yamanaka. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *cell*, 131(5):861–872, 2007.

[35] Chenyu Wu, Einar Bjarki Gunnarsson, Even Moa Myklebust, Alvaro Köhn-Luque, Dagim Shiferaw Tadele, Jorrit Martijn Enserink, Arnoldo Frigessi, Jasmine Foo, and Kevin Leder. Using birth-death processes to infer tumor subpopulation structure from live-cell imaging drug screening data. *PLoS computational biology*, 20(3):e1011888, 2024.

[36] Emily Y Yang, Grant R Howard, Amy Brock, Thomas E Yankeelov, and Guillermo Lorenzo. Mathematical characterization of population dynamics in breast cancer cells treated with doxorubicin. *Frontiers in Molecular Biosciences*, 9:972146, 2022.

[37] Liqun Yang, Pengfei Shi, Gaichao Zhao, Jie Xu, Wen Peng, Jiayi Zhang, Guanghui Zhang, Xiaowen Wang, Zhen Dong, Fei Chen, et al. Targeting cancer stem cell pathways for cancer therapy. *Signal transduction and targeted therapy*, 5(1):8, 2020.

[38] Anyue Yin, Dirk Jan AR Moes, Johan GC van Hasselt, Jesse J Swen, and Henk-Jan Guchelaar. A review of mathematical models for tumor dynamics and treatment resistance evolution of solid tumors. *CPT: pharmacometrics & systems pharmacology*, 8(10):720–737, 2019.