

# Using birth-death processes to infer tumor subpopulation structure from live-cell imaging drug screening data

Wu C.<sup>1</sup>, Gunnarsson E.B.<sup>2</sup>, Myklebust E.M.<sup>3</sup>, Köhn-Luque A.<sup>3,4</sup>, Tadele D.S.<sup>5,6</sup>,  
Enserink J.M.<sup>7,8,9</sup>, Frigessi A.<sup>3,4</sup>, Foo J.<sup>2</sup>, and Leder K.<sup>1</sup>

<sup>1</sup>Department of Industrial and Systems Engineering, University of Minnesota, Twin Cities, MN 55455, USA.

<sup>2</sup>School of Mathematics, University of Minnesota, Twin Cities, MN 55455, USA.

<sup>3</sup>Oslo Centre for Biostatistics and Epidemiology, Faculty of Medicine, University of Oslo, 0372 Oslo, Norway

<sup>4</sup>Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway

<sup>5</sup>Department of Medical Genetics, Oslo University Hospital, 0424 Oslo, Norway

<sup>6</sup>Translational Hematology and Oncology Research, Cleveland Clinic, Cleveland, OH 44131, USA

<sup>7</sup>Department of Molecular Cell Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway

<sup>8</sup>Centre for Cancer Cell Reprogramming, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Norway

<sup>9</sup>Section for Biochemistry and Molecular Biology, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway

June 14, 2023

## Abstract

Tumor heterogeneity is a complex and widely recognized trait that poses significant challenges in developing effective cancer therapies. In particular, many tumors harbor a variety of subpopulations with distinct therapeutic response characteristics. Characterizing this heterogeneity by determining the subpopulation structure within a tumor enables more precise and successful treatment strategies. In our prior work, we developed PhenoPop, a computational framework for unravelling the drug-response subpopulation structure within a tumor from bulk high-throughput drug screening data. However, the deterministic nature of the underlying models driving PhenoPop restricts the model fit and the information it can extract from the data. As an advancement, we propose a stochastic model based on the linear birth-death process to address this limitation. Our model can formulate a dynamic variance along the horizon of the experiment so that the model uses more information from the data to provide a more robust estimation. In addition, the newly proposed model can be readily adapted to situations where the experimental data exhibits a positive time correlation. We test our model on simulated data (*in silico*) and experimental data (*in vitro*), which supports our argument about its advantages.

# 1 Introduction

In recent years the design of personalized anti-cancer therapies has been greatly aided by the use of high throughput drug screens (HTDS) [21, 9]. In these studies a large panel of drugs is tested against a patient’s tumor sample to identify the most effective treatment [22, 23, 18, 3]. HTDS output observed cell viabilities after initial populations of tumor cells are exposed to each drug at a range of dose concentrations. The relative ease of performing and analyzing such large sets of simultaneous drug-response assays has been driven by technological advances in culturing patient tumor cells *in vitro*, and robotics and computer vision improvements. In principle, this information can be used to guide the choice of therapy and dosage for cancer patients, facilitating more personalized treatment strategies.

However, due to the evolutionary process by which they develop, tumors often harbor many different subpopulations with distinct drug-response characteristics by the time of diagnosis [16]. This tumor heterogeneity can confound results from HTDS since the combined signal from multiple tumor subpopulations results in a bulk drug sensitivity profile that may not reflect the true drug response characteristics of any individual cell in the tumor. Small clones of drug-resistant subpopulations may be difficult to detect in a bulk drug response profile, but these clones may be clinically significant and drive tumor recurrence after drug-sensitive populations are depleted. As a result of the complex heterogeneities present in most tumors, care must be taken in the analysis and design of HTDS to ensure that beneficial treatments result from the HTDS. In recent work we developed a method, PhenoPop, that leverages HTDS data to probe tumor heterogeneity and population substructure with respect to drug sensitivity [15]. In particular, for each drug, PhenoPop characterizes i) the number of phenotypically distinct subpopulations present, ii) the relative abundance of those subpopulations and iii) each subpopulation’s drug sensitivity. This method was validated on both experimental and simulated datasets, and applied to clinical samples from multiple myeloma patients.

In the current work, we develop novel theoretical results and computational strategies that improve PhenoPop by addressing important theoretical and practical limitations. The original PhenoPop framework was powered by an underlying deterministic population dynamic model of tumor cell growth and response to therapy. Here we introduce a more sophisticated version of PhenoPop that utilizes stochastic linear birth-death processes, which are widely used to model the dynamics of growing cellular populations [20, 14, 6], as the underlying population dynamic model powering the method. This new framework addresses several important practical limitations of the original approach: First, our original framework assumed two fixed levels of observational noise; here, the use of an underlying stochastic population dynamic model enables an improved model of observational noise that more accurately captures the characteristics of HTDS data, and reflects the observed dependence of noise amplitude on population size (see Figure 1). Second, this framework allows for natural correlations in observation noise that are tailored to fit specific experimental platforms. Rather than assuming that all HTDS observations are independent, we may consider data generated using live-cell imaging techniques where the same cellular population is studied at multiple time points, resulting in observational noise that is correlated in time. By using these stochastic processes to model the underlying populations, we obtain an improved variance and correlation structure that more accurately models the data and enables more accurate estimators with smaller confidence intervals.

The rest of the paper is organized as follows. In Section 2, we review the existing PhenoPop method and introduce the new estimation framework based on a stochastic birth-death process model of the underlying population dynamics. We propose two distinct statistical approaches in the new framework, aimed at analyzing data from endpoint vs. time series (e.g. live-cell imaging)

HTDS. In Section 3, we conduct a comprehensive investigation of our newly proposed methods and compare them with the PhenoPop method on both *in silico* and *in vitro* data. Finally, we summarize the results of the investigation and discuss the advantages of the new framework in Section 4.

## 2 Data and model formulation

The central problem we address is to infer the presence of subpopulations with different drug sensitivities using data on the drug response of bulk cellular populations. Here the term ‘bulk cellular population’ refers to the aggregate of all subpopulations within the tumor. For each given drug, we assume that the data is in the standard format of total cell counts at a specified collection of time points  $\mathcal{T} = \{t_1, \dots, t_{N_T}\}$  and drug concentrations  $\mathcal{D} = \{d_1, \dots, d_{N_D}\}$ . Furthermore, assume that for each dose-time pair,  $N_R$  independent experimental replicates are performed. We denote the observed cell count of replicate  $r$  at dose  $d$  and time  $t$  by  $x_{t,d,r}$ , and denote the total dataset by

$$\mathbf{x} = \{x_{t,d,r}; t \in \mathcal{T}, d \in \mathcal{D}, r \in \{1, \dots, N_R\}\}.$$

### 2.1 PhenoPop for drug response deconvolution in cell populations

In [15], we introduced a statistical framework for identifying the subpopulation structure of a heterogeneous tumor based on drug screen measurements of the total tumor population. Here, we briefly review the statistical framework and the resulting HTDS deconvolution method (PhenoPop). First, define the Hill equation with parameters  $(b, E, m)$  as

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

where  $b \in (0, 1)$  and  $E, m > 0$ . A homogeneous cell population treated continuously with drug dose  $d$  is assumed to grow at exponential rate  $\alpha + \log(H(d; b, E, m))$  per unit time. If the population has initial size  $C_0$ , the population size at time  $t$  is given by

$$C_0 \exp [t (\alpha + \log(H(d; b, E, m)))].$$

Note that  $H(0; b, E, m) = 1$  and  $H(d; b, E, m) \rightarrow b$  as  $d \rightarrow \infty$ . Therefore, the population grows at exponential rate  $\alpha$  in the absence of drug ( $d = 0$ ) and at rate  $\alpha + \log(b) < \alpha$  for an arbitrarily large drug dose ( $d \rightarrow \infty$ ). The parameter  $E$  represents the dose at which the drug has half the maximum effect, and  $m$  represents the steepness of the dose-response curve  $d \mapsto H(d; b, E, m)$ .

For a heterogeneous cell population, each subpopulation is assumed to follow the aforementioned growth model with subpopulation-specific parameters  $\alpha_i$  and  $(b_i, E_i, m_i)$ . Assume there are  $S$  distinct subpopulations. Then, under drug dose  $d$ , the number of cells in population  $i$  at time  $t$  is given by

$$f_i(t, d) = f_i(0) \exp [t (\alpha_i + \log(H(d; b_i, E_i, m_i)))].$$

To ease notation, the dose-response function  $H(\cdot; b_i, E_i, m_i)$  for population  $i$  will be denoted by  $H_i(\cdot)$  in what follows. The initial size of population  $i$  is  $f_i(0) = np_i$ , where  $n$  is the known initial total population size and  $p_i$  is the unknown initial fraction of population  $i$ . The total population size at time  $t$  is then given by

$$f(t, d) = \sum_{i=1}^S f_i(t, d) = n \sum_{i=1}^S p_i \exp [t (\alpha_i + \log(H_i(d)))].$$

A statistical model for the observed data  $\mathbf{x}$  is obtained by adding independent Gaussian noise to the deterministic growth model prediction. The variance of the Gaussian noise is given by

$$\sigma_{hl}^2(t, d) = \begin{cases} \sigma_H^2, & t \geq T_L \text{ and } d \leq D_L \\ \sigma_L^2, & \text{otherwise.} \end{cases}$$

The variance is allowed to depend on time and dose, since at large time points and low doses, a larger variance is expected due to larger cell counts [15]. Thus, the statistical model for the observation  $x_{t,d,r}$  is given by

$$x_{t,d,r} = f(t, d) + Z^{(r)}(t, d),$$

where  $\{Z^{(r)}(t, d); r \in \{1, \dots, N_R\}\}$  are independent random variables with the normal distribution  $N(0, \sigma_{hl}^2(t, d))$ . This model has the parameter set

$$\theta_{PP}(S) = \{(p_i, \alpha_i, b_i, E_i, m_i), \sigma_H, \sigma_L; i \in \{1, \dots, S\}\}. \quad (1)$$

The initial fractions of the  $S$  subpopulations  $\{p_i : i \in \{1, \dots, S\}\}$  and the parameters  $\{(\alpha_i, b_i, E_i, m_i) : i \in \{1, \dots, S\}\}$  governing the drug responses of the subpopulations are unknown. In addition, the variance levels  $\sigma_H^2$  and  $\sigma_L^2$  are unknown. In practice, the precise values of the thresholds  $T_L$  and  $D_L$  have minimal effect on the performance of PhenoPop. Therefore,  $T_L$  and  $D_L$  are treated as known.

The goal of the PhenoPop algorithm is to use the experimental data  $\mathbf{x}$  to estimate the unknown parameters  $\theta_{PP}(S)$  and the number of subpopulations  $S$ . The parameters  $\theta_{PP}(S)$  are estimated via maximum likelihood estimation, where the likelihood function is given by

$$L_{PP}(\theta_{PP}(S)|\mathbf{x}) = \prod_{r=1}^{N_R} \prod_{(t,d) \in \mathcal{T} \times \mathcal{D}} \frac{1}{\sqrt{2\pi\sigma_{hl}^2(t,d)}} \exp\left[-\frac{(x_{t,d,r} - f(t,d))^2}{2\sigma_{hl}^2(t,d)}\right]. \quad (2)$$

The likelihood function describes the probability of observing the data  $\mathbf{x}$  as a function of the parameter vector  $\theta_{PP}(S)$  for a given number  $S$  of subpopulations. The number of subpopulations is then estimated by comparing the negative log likelihood across candidate values of  $S$  via the elbow method or Akaike/Bayesian Information criteria. For further information, we refer to [15].

*Limitations.* The assumption of the PhenoPop algorithm that the Gaussian observation noise has two levels of variance is made for methodological simplicity and does not reflect an observed bifurcation of experimental noise levels. It would be more natural to assume that the noise level is directly proportional to the cell count, as indicated by the experimental data shown in Figure 1. In addition, PhenoPop assumes that all observations are statistically independent. However, if cells are counted using techniques such as live-cell imaging (time-lapse microscopy), then observations of the same well at different time points will be positively correlated. Both of these limitations can be addressed by modeling the cellular populations with stochastic processes, as we will now show.

## 2.2 Linear birth-death process

A natural extension of PhenoPop [15] is to use a stochastic linear birth-death process to model the cell population dynamics. In the model, a cell in subpopulation  $i$  (type- $i$  cell) divides into two cells at rate  $\beta_i \geq 0$  and dies at rate  $\nu_i \geq 0$ . This means that during a short time interval of length  $\Delta t > 0$ , a type- $i$  cell divides with probability  $\beta_i \Delta t$  and dies with probability  $\nu_i \Delta t$ . The death rate of type- $i$  cells is assumed dose-dependent according to

$$\nu_i(d) = \nu_i - \log(H_i(d)) = \nu_i - \log\left(b_i + \frac{1 - b_i}{1 + (d/E_i)^{n_i}}\right).$$

The net birth rate  $\lambda_i(d) \doteq \beta_i - \nu_i(d)$  of type- $i$  cells is then given by

$$\lambda_i(d) = (\beta_i - \nu_i) + \log(H_i(d)).$$

Using the substitution  $\alpha_i = \beta_i - \nu_i$ , we see that the drug affects the net birth rate of the stochastic model the same way it affects the growth rate  $\alpha_i$  of the deterministic population model of PhenoPop. Note however that here, the drug is assumed to act via a cytotoxic mechanism, that is, higher doses lead to higher death rates. Our framework can easily account for cytostatic effects, where higher doses lead to lower cell division rates, but we focus on cytotoxic therapies for simplicity.

Let  $X_i(t, d)$  denote the number of cells in subpopulation  $i$  at time  $t$  under drug dose  $d$ . The mean and variance of the subpopulation size at time  $t$  is given by

$$E[X_i(t, d)] \doteq np_i\mu_i(t, d) = np_i e^{\lambda_i(d)t} \quad (3)$$

$$\text{Var}[X_i(t, d)] \doteq np_i\sigma_i^2(t, d) = np_i \frac{\beta_i + \nu_i(d)}{\lambda_i(d)} \left( e^{2\lambda_i(d)t} - e^{\lambda_i(d)t} \right). \quad (4)$$

Next, denote the total population size at time  $t$  under drug dose  $d$  by

$$X(t, d) = \sum_{i=1}^S X_i(t, d),$$

with mean and variance

$$E[X(t, d)] \doteq \mu(t, d) = \sum_{i=1}^S np_i\mu_i(t, d)$$

$$\text{Var}[X(t, d)] \doteq n\sigma^2(t, d) = n \sum_{i=1}^S p_i\sigma_i^2(t, d).$$

Note that the mean size of the total population under the stochastic model equals the total population size under the deterministic model of PhenoPop, again with the substitution  $\alpha_i = \beta_i - \nu_i$ . However, the stochastic model introduces variability in the population dynamics at each time point arising from the stochastic nature of cell division and cell death. To account for experimental measurement error, we add independent Gaussian noise to each observation of the stochastic model. As a result, the new statistical model for each observation is

$$x_{t,d,r} = X^{(r)}(t, d) + Z_{t,d,r}, \quad (5)$$

where  $X^{(r)}(t, d)$  are independent copies of  $X(t, d)$  for  $r = 1, \dots, N_R$ , and  $\{Z_{t,d,r}; d \in \mathcal{D}, t \in \mathcal{T}, r \in \{1, \dots, N_R\}\}$  are i.i.d. random variables with the normal distribution  $N(0, c^2)$ , independent of the  $X^{(r)}(t, d)$ 's. The model parameter set is now

$$\theta_{BD}(S) = \{(p_i, \beta_i, \nu_i, b_i, E_i, m_i), c; i \in \{1, \dots, S\}\}. \quad (6)$$

In comparison with PhenoPop, on the one hand, the growth rate parameter  $\alpha_i$  for each subpopulation has been replaced by the birth and death rates  $\beta_i$  and  $\nu_i$ . On the other hand, there is only one parameter  $c$  for the observation noise as opposed to four parameters  $\{\sigma_H, \sigma_L, T_L, D_L\}$  for PhenoPop.

Under the new statistical model, the likelihood function is

$$L_{BD}(\theta_{BD}(S)|\mathbf{x}) = \prod_{r=1}^{N_R} \prod_{d \in \mathcal{D}} P\left(X^{(r)}(t, d) + Z_{t,d,r} \in (x_{t,d,r}, x_{t,d,r} + \Delta x_{t,d,r}), t \in \mathcal{T} | \theta_{BD}(S)\right) \quad (7)$$

where we assume that observations at different doses and from distinct replicates are independent, and  $(x_{t,d,r}, x_{t,d,r} + \Delta x_{t,d,r})$  represents an infinitesimally small interval around  $x_{t,d,r}$ . We now discuss two different forms this likelihood function can take, depending on whether the data collected at different time points are correlated or not.

### 2.2.1 End-point experiments

For many common cell counting techniques, e.g. CellTiter-Glo [13], the experiment must be stopped to perform the viability assay. In this case, observations at different time points are actually observations of different cell populations exposed to drugs for different amounts of time, and can therefore be treated as independent. Thus, the likelihood function can be written as

$$L_{EP}(\theta_{BD}(S)|\mathbf{x}) = \prod_{r=1}^{N_R} \prod_{d \in \mathcal{D}} \prod_{t \in \mathcal{T}} P\left(X^{(r)}(t, d) + Z_{t,d,r} \in (x_{t,d,r}, x_{t,d,r} + \Delta x_{t,d,r}) | \theta_{BD}(S)\right).$$

We note that the distribution of  $X^{(r)}(t, d) + Z_{t,d,r}$  can be computed exactly. However, for faster computation, one can approximate the distribution by a Gaussian distribution. To that end, consider the centered and normalized process

$$W_n(t, d) = \frac{1}{\sqrt{n}} \sum_{i=1}^S \left( X_i(t, d) - np_i e^{\lambda_i(d)t} \right). \quad (8)$$

A straightforward application of the central limit theorem gives the following result.

**Proposition 1** For  $t > 0$  and  $d \geq 0$ ,  $W_n(t, d) \Rightarrow N(0, \sigma^2(t, d))$ , as  $n \rightarrow \infty$ .

Note that ‘ $\Rightarrow$ ’ means converge in distribution. The proof of this result will not be provided since it is a consequence of the more general Proposition 2.

Based on Proposition 1, we obtain the likelihood function

$$L_{EP}(\theta_{BD}(S)|\mathbf{x}) = \prod_{r=1}^{N_R} \prod_{(t,d) \in \mathcal{T} \times \mathcal{D}} \frac{1}{\sqrt{2\pi(n\sigma^2(t, d) + c^2)}} \exp \left[ -\frac{(x_{t,d,r} - \mu(t, d))^2}{2(n\sigma^2(t, d) + c^2)} \right]. \quad (9)$$

The right-hand side of (9) depends on the model parameters  $\theta_{BD}(S)$  via the mean and variance functions  $\mu(t, d)$  and  $\sigma^2(t, d)$ . As in [15], one can maximize this expression over the parameter set  $\theta_{BD}(S)$  to obtain maximum likelihood estimates of the model parameters. The optimization problem for the new likelihood  $L_{EP}$  is more difficult to solve than the corresponding problem for the PhenoPop likelihood  $L_{PP}$  in (2), since the variance of the data now depends on the dose-response parameters for the subpopulations. However, the numerical optimization software we employ is able to deal with this more complex dependence on the model parameters, see Appendix 5.1.

### 2.2.2 Live-cell imaging techniques

Live-cell imaging techniques enable the experimenter to obtain cell counts for the same population across multiple different time points. For such datasets, observations of the same sample at different time points will be positively correlated. In this case, we must compute the joint distribution

$$P\left(X^{(r)}(t, d) + Z_{t,d,r} \in (x_{t,d,r}, x_{t,d,r} + \Delta x_{t,d,r}), t \in \mathcal{T} | \theta_{BD}(S)\right) \quad (10)$$

for each  $d \in \mathcal{D}$ . To ease notation, we will temporarily suppress dependence on the dose.

We first note that  $(X^{(r)}(t))_{t \geq 0}$  is not a Markov process, since the total cell count at each time point does not include information on the sizes of the individual subpopulations. Computing (10) exactly requires summing over the possible sizes of the subpopulations at each time point, which is computationally intensive. It is possible to speed up the computation using tools from hidden Markov models, which reduces the computational complexity to  $\Omega(\min_{t \in \mathcal{T}} x_t^2)$ . However, this is still computationally infeasible since  $\min_{t \in \mathcal{T}} x_t \approx 5000$ , resulting in a 1 second computation time to evaluate a single likelihood. The computational details are provided in Appendix 5.5.

A more efficient approach is to use a Gaussian approximation. For the centered and normalized process  $W_n(t)$  from (8), define the vector of observations across time points

$$\mathbf{W}_n = \{W_n(t); t \in \mathcal{T}\}.$$

By assuming that the set  $\mathcal{T}$ , number of subpopulations  $S$  and initial proportion  $p_i$  of each subtype  $i$  are independent of the initial total cell count  $n$ , we derive the following approximation for  $\mathbf{W}_n$ :

**Proposition 2** As  $n \rightarrow \infty$ ,

$$\mathbf{W}_n \Rightarrow \mathbf{Y} = \{Y(t); t \in \mathcal{T}\} \sim N(0, \Sigma),$$

where the  $(i, j)$  element of the covariance matrix  $\Sigma$  is given by

$$\Sigma_{i,j} = \sum_{\ell=1}^{\min(i,j)} \sum_{k=1}^S p_k e^{(\lambda_k t_i - \lambda_k t_\ell)} e^{(\lambda_k t_j - \lambda_k t_\ell)} e^{\lambda_k t_{\ell-1}} \sigma_k^2 (t_\ell - t_{\ell-1}).$$

The proof of this proposition is given in Appendix 5.3. In Appendix 5.4, we relax the assumption that the initial proportion  $p_i$  is independent of the initial total cell count, and a similar result still follows. In future work, we plan to relax the assumption that  $\mathcal{T}$  is independent of  $n$ .

We now reintroduce dose dependence. For each  $d \in \mathcal{D}$ , define the  $N_T \times 1$  vector

$$\mu(d) = \{\mu(t, d); t \in \mathcal{T}\}$$

and the  $N_T \times N_T$  identity matrix  $I$ . Based on Proposition 2, the following approximation is used to compute the likelihood in expression (7):

$$\mathbf{x}_{\cdot, d, r} = (x_{t, d, r}; t \in \mathcal{T}) \approx \mu(d) + N(0, n\Sigma(d)) + N(0, c^2 I).$$

The likelihood function is thus given by

$$L_{LC}(\theta_{BD}(S) | \mathbf{x}) = \prod_{r=1}^{N_R} \prod_{d \in \mathcal{D}} \frac{\exp \left[ -\frac{1}{2} (\mathbf{x}_{\cdot, d, r} - \mu(d))^T (n\Sigma(d) + c^2 I)^{-1} (\mathbf{x}_{\cdot, d, r} - \mu(d)) \right]}{(\det(2\pi(n\Sigma(d) + c^2 I)))^{1/2}}. \quad (11)$$

Note that the computational complexity of evaluating the above likelihood is independent of  $\min_{t \in \mathcal{T}} x_t$ , alleviating the computational burden associated with an exact evaluation of the likelihood.

The difference between the likelihood function  $L_{EP}$  for endpoint data and  $L_{LC}$  for live-cell imaging data lies in the structure of the covariance matrix for the observation vector  $\mathbf{x}_{\cdot, d, r}$ . For  $L_{EP}$ , observations made at different time points are assumed independent, meaning that the covariance matrix is diagonal. For live-cell imaging data, the covariance matrix is not diagonal. Accurately accounting for time correlations in the likelihood (11) can improve the accuracy of parameter estimates, as we will discuss in Section 3. However, it does come at a cost, since it is obviously more computationally expensive to calculate the inverses and determinants present in  $L_{LC}$ . As a result, the optimization of  $L_{LC}$  can be more difficult than the optimization of  $L_{EP}$ .

### 2.2.3 Accuracy of Gaussian approximation

Proposition 2 states that the centered and normalized process  $\mathbf{W}_n$  is approximately Gaussian  $N(0, \Sigma(d))$ . However, in the derivation of the likelihood function (11), the distribution of the total cell number  $\mathbf{X}(d) = \{X(t, d); t \in \mathcal{T}\}$  is approximated with a Gaussian distribution  $N(\mu(d), n\Sigma(d) + c^2I)$ , whose mean and variance increases linearly with  $n$ . To verify that the error in this approximation is reasonable for large  $n$ , we will now compare the distributions of  $\mathbf{X}(d)$  and  $N(\mu(d), n\Sigma(d) + c^2I)$  using a well-known measure of the distance between two distributions.

The energy distance, introduced in [24], is a measure of the distance between probability distributions, which has previously been shown to be related to Cramer’s distance [5, 24]. The energy distance has been utilized in several statistical tests [2] and is easily computed for multivariate distributions. For probability distributions  $F$  and  $G$  on  $\mathbb{R}^d$ , we define their energy distance as

$$D(F, G) = \sqrt{2\mathbb{E}[\|X - Y\|] - \mathbb{E}[\|X - X'\|] - \mathbb{E}[\|Y - Y'\|]}, \quad (12)$$

where all random variables are independent,  $X$  and  $X'$ , and  $Y$  and  $Y'$ , are distributed according to  $F$  and  $G$  respectively, and  $\|\cdot\|$  denotes the Euclidean norm.

Since it is unrealistic to compute the equation (12) directly, we approximate the true energy distance by computing the empirical energy distance. For two sets of i.i.d. realization  $\{X_1, \dots, X_k\}, X_i \sim F, \{Y_1, \dots, Y_m\}, Y_i \sim G$ , one can obtain the empirical energy distance by

$$D_E(F, G) = \sqrt{\frac{2}{km} \sum_{i=1}^k \sum_{j=1}^m \|X_i - Y_j\| - \frac{1}{k^2} \sum_{i=1}^k \sum_{j=1}^k \|X_i - X_j\| - \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \|Y_i - Y_j\|}. \quad (13)$$

Denote the distribution of  $\mathbf{X}(d)$  by  $F_{BD}$  and the normal distribution  $N(\mu(d), n\Sigma(d) + c^2I)$  by  $F_N$ . Let  $\{X_i\}_{i=1}^k$  be  $k$  i.i.d. samples from the distribution  $F_{BD}$ , and let  $\{Y_i\}_{i=1}^m$  be  $m$  i.i.d. samples from  $F_N$ . We can then compute  $D_E(F_{BD}, F_N)$  using (13). In Figure 2, we plot  $D_E(F_{BD}, F_N)$  with varying initial cell counts. The plot shows a monotonic decrease in the empirical energy distance as a function of the initial cell count, which indicates that the distribution of  $\mathbf{X}(d)$  is reasonably approximated by a Gaussian distribution for large values of the initial cell count.

## 3 Numerical results

In this section, we use our new statistical methods to analyze both simulated (*in silico*) and experimental (*in vitro*) live cell imaging data. We apply both the simpler end-point estimation procedure (“end-points method”), based on the likelihood  $L_{EP}$  in (9), and the more complex live cell imaging procedure (“live cell image method”), based on the likelihood  $L_{LC}$  in (11). The performance of the new methods is compared with the existing PhenoPop algorithm. In all analyses it is assumed that the observation at time  $t = 0$  represents the known starting population size, i.e.  $x_{0,d,r} = n$ .

### 3.1 Application to simulated data

We first apply our estimation methods to simulated (*in silico*) data. In Appendix 5.1, we provide details of the data generation and the parameter estimation for these *in silico* experiments.

#### 3.1.1 Examples with 2 subpopulations

For illustrative purposes, we begin with a case study involving an artificial tumor with two subpopulations. Data is generated using a parameter vector  $\theta_{BD}(2)$  selected uniformly at random



from the ranges in Table 1. We assume that one tumor subpopulation is drug-sensitive and the other is drug-resistant. These subpopulations are indicated by the subscripts  $s$  and  $r$ , respectively.

	$p_s$	$p_r$	$\beta_{s,r}$	$\nu_{s,r}$	$b_{s,r}$	$E_s$	$E_r$	$m_{s,r}$	$c$
Range	[0.3, 0.5]	$1 - p_s$	[0, 1]	$[\beta - 0.1, \beta]$	[0.8, 0.9]	[0.05, 0.1]	[0.75, 2.5]	[1.5, 5]	[0, 10]

Table 1: Range for parameter generation of experiments with 2 subpopulations

As in [15], we focus on inferring the initial proportion  $p_s$  of sensitive cells, as well as the  $GR_{50}$  dose for each subpopulation. The  $GR_{50}$  is the dose at which the drug has half the maximal effect on the cell death rate, as is further explained in Appendix 5.1. Informally, the  $GR_{50}$  dose for each subpopulation is a measure of the subpopulation’s sensitivity to the drug. To assess the uncertainty in the parameter estimation, we compute maximum likelihood estimates for 100 bootstrapped datasets, as described in Appendix 5.1. The results of the case study are shown in Figure 3, where we see that both the live cell image method and the end-points method are able to recover the initial proportion  $p_s$  of the sensitive population and the  $GR_{50}$  dose for each subpopulation accurately.

We next evaluate the performance of the estimation methods across 30 simulated datasets, where each parameter vector  $\theta_{BD}^i(2)$  for  $i = 1, \dots, 30$  is sampled from the ranges in Table 1. We furthermore compare the performance of the two new methods with the performance of PhenoPop. The error in the estimation of each parameter  $\{p_s, GR_s, GR_r\}$  is measured by considering the absolute log ratio between the point estimate  $\hat{x}$  and the true value  $x$  for the parameter,

$$Er(\hat{x}; x) = \left| \log \left( \frac{x}{\hat{x}} \right) \right|. \quad (14)$$

This metric is chosen to address the logarithmic scale associated with the  $GR_{50}$  dose.

In Figure 4, a box plot of the estimation errors for the three methods across the 30 datasets is presented. Note that all three parameters  $\{p_s, GR_s, GR_r\}$  are estimated accurately using all three methods. In addition, the error in estimating the sensitive  $GR_{50}$  is larger than the error in estimating the resistant  $GR_{50}$  for all three methods. One possible reason is that the initial proportion of sensitive cells is  $p_s \in [0.3, 0.5]$ , so the experimental data contains less information on the sensitive subpopulation. Later experimental results will lend further support to this hypothesis.

We next compare the estimation precision of the three methods. Specifically, we will compare the widths of the 95% confidence intervals for the three parameters between the three different methods. Since  $GR_{50}$  values vary significantly across the 30 generated datasets, we normalize the CI width for each method by dividing it by the sum of the CI widths of all three methods for the same dataset.

In Figure 5, the normalized CIs for  $\{p_s, GR_s, GR_r\}$  are compared between the three methods across the 30 datasets. First, note that for the initial proportion  $p_s$ , the live cell image method has significantly narrower CIs than the other two methods. Additionally, there is a small but statistically significant difference between the CI widths for the end-points method and the PhenoPop method. For the sensitive  $GR_{50}$  index, the live cell image method again has significantly narrower confidence intervals than the other two methods, and the end-points method has significantly narrower confidence intervals than the PhenoPop method. The results are similar for the resistant  $GR_{50}$  index. It is worth mentioning that for at least 28 out of the 30 datasets, the true parameters were located within the confidence intervals for all three methods.

In summary, the end-points and live cell image methods provide a significant improvement in estimator precision over the PhenoPop method for all three parameters, and furthermore, the live cell image method has the best precision out of all three methods.

### 3.1.2 Illustrative example with 3 subpopulations

In this section, we examine a case study involving an artificial tumor with 3 subpopulations. The subpopulations are assumed sensitive, moderate, and resistant with respect to the drug, and they are denoted using the subscripts  $s$ ,  $m$ , and  $r$ , respectively. Data is generated using a parameter vector  $\theta_{BD}(3)$  selected uniformly at random from the ranges in Table 2. Those parameters not listed in Table 2 are selected as in Table 1.

	$p_s, p_m$	$p_r$	$E_s$	$E_m$	$E_r$
Range	[0.167, 0.333]	$1 - p_s - p_m$	[0.0313, 0.0625]	[0.25, 0.375]	[1.25, 2.5]

Table 2: modified range of parameters

Figure 6 shows estimation results for the initial proportions  $p_s, p_m, p_r$  and the  $GR_{50}$  doses of the three subpopulations. Note that the end-points and live cell image methods provide more accurate estimates of the initial proportion for each subpopulation than PhenoPop. Furthermore, when estimating the  $GR_{50}$  for each subpopulation, the inter-quartile range (IQR) of 100 bootstrapped estimates covers the true  $GR_{50}$  value for all three methods. However, the estimation for the  $GR_{50}$  of the moderate subpopulation with  $E_m = 0.3558$  is less precise than for the other two subpopulations, i.e., the IQR is wider. This is likely due to confounding between the moderate subpopulation and the other two subpopulations.

It is worth noting that for the 3 subpopulation example, the number of datapoints is the same as for the 2 subpopulation examples, since only total cell counts are observed at each time point. Furthermore, when computing maximum likelihood estimates for 3 subpopulations, we solved each optimization problem the same number of times as for 2 subpopulations. Overall, our conclusion is that all three methods can provide reasonable estimates of the true initial proportion and the  $GR_{50}$  of each subpopulation for 3 subpopulations. However, achieving equivalent levels of accuracy and precision as for 2 subpopulations may require a greater computational effort or the collection of more data, given that the 3 subpopulation model is more complex and has more parameters.

### 3.1.3 Performance in challenging conditions

In the previous work [15], three conditions under which the performance of PhenoPop deteriorates were identified: the case of a large observation noise, a small initial fraction of resistant cells, and similar drug-sensitivity of both subpopulations. We now investigate the performance of the end-points and live cell image methods in these conditions and compare to the performance of PhenoPop.

#### Large observation noise:

We first consider the case of large observation noise. Note that in the PhenoPop method, the only source of variability in the statistical model is the additive Gaussian noise. In the end-points and live cell image methods, however, there is an underlying stochastic process governing the population dynamics with an added Gaussian noise term. Thus, whereas PhenoPop deals with high levels of noise by adjusting the variance of the Gaussian term, the two new methods may also try to adjust the subpopulation growth and dose response parameters. This can complicate estimation with the two new methods compared to PhenoPop from data with high levels of noise.

We begin by considering a case study where the noise level is set to  $c = 500$ , and other parameters are chosen uniformly at random according to Table 1. The results are shown in Figure 7. For each method, the initial proportion  $p_s$  is estimated with good accuracy, and the IQR of 100 bootstrap

estimates for  $GR_s$  covers the true value. However, compared with the estimation in Figure 3, the estimation precision of the end-points method and live cell image method has degraded. In addition, observe that the IQRs of the three methods have about the same width, which implies the precision advantage observed in Section 3.1.1 disappears under a very large observation noise.

We next evaluate estimation performance across 30 simulated datasets for each noise value  $c \in \mathcal{C} = \{100, 200, 300, 400, 500\}$ . Figure 8 shows the mean absolute log ratio across the 30 datasets for each parameter  $\{p_s, GR_s, GR_r\}$ , each noise level and each estimation method. As expected, the estimation error increases for all three methods as a function of the observation noise. In fact, all three methods show a similar response to increasing levels of noise.

We next compare the widths of 95% confidence intervals for the three parameters under noise levels  $c = 100$  and  $c = 500$ , using 30 datasets for each noise level. The results are shown in Figures 9 and 10. For  $c = 100$  (Figure 9), the precision advantage of the live cell image method over the other two methods is less pronounced than in Figure 5, where  $c \in [0, 10]$ , especially for the resistant  $GR_{50}$ . For  $c = 500$  (Figure 10), the advantage disappears for the sensitive  $GR_{50}$ . Importantly, however, Figure 9 shows that the precision advantage of the live cell method is statistically significant for all three parameters  $\{p_s, GR_s, GR_r\}$  for an observation noise as large as 10% of the initial cell count. It should be noted that the standard deviation of observation noise reported from common automated and semi-automated cell counting techniques ranges from 1 – 15% [4, 19].

### Small resistant subpopulation:

For the datasets investigated in Section 3.1.1, the initial proportion  $p_s$  of sensitive cells was constrained to be in  $[0.3, 0.5]$ . We now consider the setting of a small resistant subpopulation. We begin with a case study in Figure 11, where  $p_s$  is assigned to 0.99, and other parameters are sampled according to Table 1. For both the sensitive and resistant subpopulations, the IQR for the  $GR_{50}$  dose under PhenoPop does not cover the true value, whereas the IQR for the live cell image method does. The IQR for the end-points method covers the true resistant  $GR_{50}$ , but only barely covers the true sensitive  $GR_{50}$ . In addition, the end-points and live cell image methods have significantly narrower IQRs than PhenoPop. Finally, note that the estimate of the initial proportion of resistant cells is much more accurate for the end-points and live cell image methods. Thus, while PhenoPop provides a reasonable estimate of the sensitive  $GR_{50}$ , which is the dominant subpopulation in this scenario, inferring the population composition and the  $GR_{50}$  for the minority resistant subpopulation requires the use of the more powerful end-points and live cell image methods.

In Figure 12, we show the mean absolute log ratio for each parameter  $\{p_s, GR_s, GR_r\}$  across 100 datasets for each  $p_s \in \{0.85, 0.9, 0.95, 0.99\}$ . Note that both the end-points and live cell image methods have significantly smaller errors than PhenoPop, and that the difference becomes more pronounced as  $p_s$  increases. Also note that the error in estimating the sensitive  $GR_{50}$  is smaller than for the resistant  $GR_{50}$ , opposite to the results of Figure 4, where  $p_s \in [0.3, 0.5]$ . This further reinforces the hypothesis stated in Section 3.1.1 that the initial proportion of a subpopulation impacts the precision of estimating the  $GR_{50}$  for that subpopulation.

### Similar subpopulation sensitivity:

For the datasets investigated in Section 3.1.1, the  $GR_{50}$ 's for the two subpopulations were assumed to be significantly different. We now consider the case where the two  $GR_{50}$ 's are similar. Figure 13 shows the results of a case study where  $E_s \in [0.05, 0.1]$ ,  $E_r = 0.15$ , and other parameters are selected according to Table 1. Note that all three methods successfully recover the parameters  $\{p_s, GR_s, GR_r\}$ , where the IQRs for the live cell image method are significantly

narrower than for PhenoPop. For brevity, we omit the plots that depict the statistical comparison of confidence interval widths. In Figure 14, we perform estimation across 80 datasets for each  $E_r \in \{0.15, 0.3, 0.45, 0.85, 2.0\}$ , with other parameters sampled from Table 1, including  $E_s \in [0.05, 0.1]$ . As expected, the accuracy in estimating the parameters  $\{p_s, GR_s, GR_r\}$  improves as the sensitive  $GR_{50}$  and resistant  $GR_{50}$  become more different. We note however that the live cell image method has the lowest mean error when estimating the parameters, with all three methods showing similar degradation as the two subpopulations become more phenotypically similar.

### 3.2 Application to *in vitro* data

We conclude by evaluating the performance of our two new methods on *in vitro* experimental data. The data consists of different mixtures of imatinib sensitive and resistant *Ba/F3* cells. In the experiments, cells were exposed to 11 different concentrations of imatinib and they were observed at 14 different time points. For each drug concentration, 14 independent replicates were performed starting with roughly 1000 cells. Cell counts were obtained using a live-cell imaging technique. Four datasets were produced with different starting ratios between sensitive and resistant cells: 1 : 1, 1 : 2, 2 : 1 and 4 : 1. These datasets are denoted BF11, BF12, BF21 and BF41, respectively. See [15] for further details on the experimental methods for generating the data.

In [15], we showed that the PhenoPop method can accurately identify the initial proportion of sensitive cells and both subpopulations'  $GR_{50}$  indices from the datasets. Here, we apply the two new estimation methods to the datasets and compare how well the models fit the data. Model fits are assessed using the Akaike Information Criterion (AIC), which for a statistical model with parameters  $\theta$  and likelihood function  $\mathcal{L}(\theta|\mathbf{x})$  is given by

$$AIC = -2\log(\mathcal{L}(\theta^*|\mathbf{x})) + 2|\theta^*|.$$

Here,  $\theta^*$  is the maximum likelihood estimate and  $|v|$  is the cardinality of the vector  $v$ . When comparing the three methods, the one with the lowest AIC is preferred.

DATA	PP(AIC)	EP(AIC)	LC(AIC)
BF11	28502	26300	<b>25294</b>
BF12	30485	<b>26816</b>	27311
BF21	27928	<b>24064</b>	24182
BF41	28912	<b>24066</b>	24574

Table 3: AIC scores of three methods: PhenoPop method(PP), end-points method(EP), and live cell image method(LC) for the four experimental datasets BF11, BF12, BF21 and BF41.

Results are shown in Table 3. The AIC values of the end-points method (EP) and live cell image method (LC) are clearly lower than for the PhenoPop method (PP), indicating that the two new methods are superior for fitting the experimental datasets. As discussed in Section 2, the newly proposed methods have more sophisticated variance structures, which is likely the reason why they are able to provide a better fit to the datasets. Finally, we note that the end-points method has superior AIC scores to the live cell image method for three out of four of the experimental conditions. Therefore, it is not clear which of these two methods is more appropriate for fitting these datasets.

## 4 Discussion

In this work, we have proposed two methods for analyzing data from heterogeneous cell mixtures. In particular, we are interested in the setting where a mixture of at least two distinct cell subpopulations is exposed to a given drug at various concentrations. We then use the dose response curve of the composite population to learn about the two subpopulations. In particular, we are interested in estimates of the different subpopulations' initial prevalence and also their distinct dose response curves. The challenge of this problem is that we do not observe direct information about the subpopulations, but instead only information about the dose response of the composite population.

This work is an extension of our prior work in [15]. The novelty of the current work is that we introduce a more realistic variance structure to our statistical model. We create a new variance structure by building our model using linear birth-death processes. In particular, we model each subpopulation as a linear birth-death process with a unique birth rate and a unique dose-dependent death rate. The dose dependence of the death rate is captured using a 3-parameter Hill function. Our observed process is then a sum of independent birth-death processes. Our goal is then to estimate the initial proportion of the subpopulations, as well as their birth rates and the parameters governing the dose response in their death rates.

Counting cells in *in vitro* experiments can generally be conducted in one of two fashions. In the first approach, cell numbers can only be estimated at the end of the experiment because the mechanism for estimating cell numbers requires killing the cells. In the second approach, cells are counted via live imaging techniques and the cells can be counted at multiple time points. When dealing with multiple time point data from cells collected via the first approach we can assume that observations at different time points are independent because they are the result of different experiments. However, when dealing with data from the second approach we can no longer make that assumption because the cell counts at different time points are from the same population and there is a positive correlation between those measurements. As a result of this differing structure we develop two methods, one that assumes independent observations at each time point, and one that assumes all the time points for a given dose are correlated. Evaluating the likelihood function under the second approach is not trivial at first glance since it requires evaluating the likelihood function of a sample path of a non-Markovian process (the total cell count). We are able to get around this difficulty by using a central limit theorem argument to approximate the exact likelihood function with a Gaussian likelihood.

In this work we compared three different methods: PhenoPop method from [15], end-points method (assumes measurements are independent in time), and live cell image method (assumes time correlations). We first performed this comparison using simulated data. We generated our data by simulating linear birth-death processes and then adding independent Gaussian noise terms to the simulations. We mainly focused on a mixture of two subpopulations, and we were interested in estimating three features of the mixed population: initial proportion of sensitive cells,  $GR_{50}$  of the sensitive cells, and  $GR_{50}$  of the resistant cells. Our first test for the simulated data was to look at confidence interval widths as a measure of estimator precision. In this study, we found that the live cell image method had significantly narrower confidence intervals than the other methods for estimating all three features. We next investigated the performance of our three methods in the setting of small resistant subpopulations, where less than 15% of initial cells are resistant. We found that in this small resistant fraction setting the live cell image method provides a significant improvement in accuracy over the original PhenoPop method. Furthermore, this improvement increases as the initial fraction of resistant cells goes to zero. We also compared the performance of the methods for simulations with increased levels of additive noise and subpopulations with similar dose response curves. In the scenario of subpopulations with similar dose response curves, we found that the live

cell image method has the lowest mean error among the three methods. For increasing additive noise, all three methods perform similarly in terms of estimation accuracy. However, the live cell image method maintains its precision advantage over the other two methods for an observation noise of 10% of the initial cell count, while the advantage disappears for a 50% noise level.

We finally compared the three methods using *in vitro* data. In particular, we used data from our previous work [15] that considered different seeding mixtures of imatinib sensitive and resistant tumor cells. We then used all three methods to fit this data and used AIC as a model selection tool. We found that live cell image and end-points methods had significantly better scores than PhenoPop for all four initial mixtures studied. Interestingly the end-points method had lower AIC scores for three out of the four mixtures studied even though this data was generated using live-cell imaging techniques.

In our statistical model, there are several important features of cell biology that we have left out. For example, one type of cell may transition to another type of cell via a phenotypic switching mechanism (see e.g., [10, 11]). We believe that our current methods should be able to handle this type of switching with little modification since the underlying stochastic model will be very similar, i.e., a multi-type branching process. Another way the cell types can interact is via competition for scarce resources as the populations approach their carrying capacity. These types of interactions will require new statistical models since the underlying stochastic processes will no longer be linear birth-death processes. Another interesting direction of future work is to quantify the limits of when we can identify distinct subpopulations. For example, if the resistant subpopulation is present at fraction  $\epsilon$ , what observation set would allow us to identify the presence of this subpopulation? Finally our stochastic model assumes that the time between cell divisions is exponential, but this is of course a great simplification. At the cost of a more complex model it would be possible to incorporate states for the different stages of the cell cycle. We leave this open as a question for future investigation.

## 5 Appendix

### 5.1 Details of the numerical experiments

We define some of the basic algorithms and formulas used in the numerical results.

#### Generation of simulated data

To simulate data, the parameter set  $\theta_{BD}(S)$  (as defined in (6)) is selected uniformly at random from a subset of the parameter space given in Table 1. Note that one can obtain the generating parameter set  $\theta_{PP}(S)$  (defined in (1)) from  $\theta_{BD}(S)$  directly by setting  $\alpha_i = \beta_i - \nu_i$  for each subpopulation. Based on the parameter set  $\theta_{BD}(S)$ , we simulated data is generated according to the statistical model specified in equation (5). Note that data is collected from the simulation continuously during the course of the experiment to replicate the live-cell imaging experiments.

#### Maximum likelihood estimation (MLE)

The maximum likelihood estimation was conducted by minimizing the negative log-likelihood, subject to constraints that were placed on the range of each parameter. The optimization process to find the minimum point was based on the MATLAB Optimization Toolbox [17] function `fmincon` with sequential quadratic programming (sqp) solver. Due to the non-convexity of the negative log-likelihood function, we performed the optimization starting from 100 uniformly sampled initial points within a feasible region. The feasible region sets limitations on the parameters based on prior

knowledge about them. For simulation studies, the feasible region is given by Table 4, and for the *in vitro* data the feasible region is specified by Table 5. Among all the resulting local optima, the parameter set with the lowest negative log-likelihood as the estimated result.

### Bootstrapping

In the simulated experiments, bootstrapping is used to quantify the uncertainty in the MLE estimator. In particular, 20 independent replicates of data measured at 11 concentration values  $\mathcal{D}$  and 13 time points  $\mathcal{T}$  are generated from the parameters  $\theta_{BD}(S)$  at the beginning of the experiment. Then bootstrapping is employed to randomly re-sample 13 replicates from those 20 replicates with replacement 100 times. With 13 randomly sampled replicates it is possible to create an MLE for the parameter set  $\theta_{BD}$ . Since there are now 100 MLE's for  $\theta_{BD}$  it is possible to construct confidence intervals as well by using the empirical quantiles of the estimators.

### $GR_{50}$

Our goal is to estimate the number of subpopulations, initial mixture proportion  $p_i$ , and the drug sensitivity of each cellular subpopulation. The  $GR_{50}$ , introduced in [12], is a 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. Note that at the maximum concentration level, the drug may not reach its theoretical maximum effect.

In the context of the model, the  $GR_{50}$  can be defined as below. Denote the maximum dosage applied as  $d_m$ , and define the half-maximum effect for subpopulation  $i$  as  $r_i = (\nu_i(d_m) + \nu_i(0))/2$ . The explicit formula for the  $GR_{50}$  is then for subpopulation  $i$ :

$$GR_{50} = E_i \left( \frac{1 - e^{\nu_i - r_i}}{e^{\nu_i - r_i} - b_i} \right)^{1/m_i}$$

When  $S = 2$ , we will denote the higher  $GR_{50}$  as either the resistant  $GR_{50}$  or  $GR_r$ , and the lower  $GR_{50}$  as either the sensitive  $GR_{50}$  or  $GR_s$ . In this setting, parameters for the sensitive subpopulation and the resistant subpopulation, respectively, are denoted by subscripts  $s$  and  $r$ , e.g.  $E_s$  and  $E_r$ .

### Initial conditions.

The initial number of cells is set as  $n = 1000$ , and the initial size of each subpopulation is set by rounding  $np_i$  to the nearest integer for subpopulation  $i$ . The following drug concentration levels are used

$$\mathcal{D} = [0, 0.0313, 0.0625, 0.1250, 0.2500, 0.3750, 0.5, 1.25, 2.5, 3.75, 5]$$

and we collect the cell count data at the time points:

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

For these specific concentration levels and time points, we have chosen the threshold values of  $T_L = 21$  and  $D_L = 1$  in the PhenoPop model.

### Optimization feasible region

When performing numerical optimization the parameters are restricted to a physically realistic region. Unless otherwise noted, the optimization was performed using 100 uniformly sampled initial points from Table 4.

	$p_s$	$p_r$	$\beta_{s,r}$	$\nu_{s,r}$	$b_{s,r}$	$E_{s,r}$	$m_{s,r}$	$\sigma_L, \sigma_H$	$c$
Range	[0, 0.5]	$1 - p_s$	[0, 1]	$[\beta - 0.1, \beta]$	[0.27, 1]	[0, 10]	[0, 10]	[0, 2500]	[0, 10]

Table 4: Feasible interval for each parameter.

## 5.2 Estimation on *in vitro* experimental data

When solving the maximum likelihood optimization problems for the *in vitro* data of Section 3.2, the optimization feasible region was chosen to be the same as the feasible region used in paper [15] for the *Ba/F3* data, i.e.,

	$p$	$\beta$	$\nu$	$b$	$E$	$m$	$\sigma_L, \sigma_H$	$c$
Range	[0, 1]	[0, 1]	$[\beta - 0.06, \beta]$	[0.878, 1]	[0, 50]	[0.001, 20]	[0, 2500]	[0, 100].

Table 5: Optimization Feasible region

We solved each optimization problem 500 times starting from randomly chosen initial points.

## 5.3 Proof of proposition 2

Given a set of time points  $\mathcal{T} = \{t_1, \dots, t_k\}$ , we first show for any  $t_j, 1 \leq j \leq k$  that:

$$W_n(t_j) \Rightarrow Y(t_j) = \sum_{\ell=1}^j \sum_{g=1}^S \sqrt{p_g} e^{\lambda_g(t_j - t_\ell)} e^{\lambda_g t_{\ell-1}/2} V_g(t_\ell - t_{\ell-1}) \text{ as } n \rightarrow \infty$$

Where  $V_g(t)$  is a random variable that has normal distribution  $N(0, \sigma_g^2(t))$  and the  $\sigma_g^2(t)$  here is the variance of subpopulation  $g$  linear birth-death process defined in equation (4). Following an argument from Either and Kurtz [7], we have the decomposition

$$\begin{aligned} W_n(t_j) &= \frac{1}{\sqrt{n}} \sum_{g=1}^S \sum_{\ell=1}^j e^{\lambda_g(t_j - t_\ell)} X_g(t_\ell) - e^{\lambda_g(t_j - t_{\ell-1})} X_g(t_{\ell-1}) \\ &= \sum_{\ell=1}^j \sum_{g=1}^S \frac{1}{\sqrt{n}} e^{\lambda_g(t_j - t_\ell)} \left[ X_g(t_\ell) - e^{\lambda_g(t_\ell - t_{\ell-1})} X_g(t_{\ell-1}) \right] \\ &= \sum_{\ell=1}^j \sum_{g=1}^S e^{\lambda_g(t_j - t_\ell)} \left( \frac{p_g n}{n} \right)^{1/2} \left( \frac{X_g(t_{\ell-1})}{p_g n} \right)^{1/2} X_g(t_{\ell-1})^{-1/2} \sum_{m=1}^{X_g(t_{\ell-1})} \left[ B_g(t_\ell - t_{\ell-1}) - e^{\lambda_g(t_\ell - t_{\ell-1})} \right] \end{aligned}$$

By assuming the initial proportion  $p_g$  for sub-type  $g$  is independent of  $n$  and using the Law of large numbers, as  $n \rightarrow \infty$ , we have

$$\left( \frac{X_g(t_{\ell-1})}{p_g n} \right) \rightarrow \mathbb{E}[B_g(t_{\ell-1})] = e^{\lambda_g t_{\ell-1}} \quad a.s.$$

By assuming that the maximum number of time point  $N_T$  and the length of time interval  $t_i - t_j$  for any  $i \geq j$  are both bounded and not depend on  $n$ , the Law of large numbers also assures that for



any  $\ell \in \{1, \dots, N_T\}$  the  $X_g(t_{\ell-1})$  will diverge to infinity when  $n \rightarrow \infty$ . Therefore, we may apply the Central Limit Theorem to the following term:

$$X_g(t_{\ell-1})^{-1/2} \sum_{m=1}^{X_g(t_{\ell-1})} \left[ B_g(t_\ell - t_{\ell-1}) - e^{\lambda_g(t_\ell - t_{\ell-1})} \right] \Rightarrow V_g(t_\ell - t_{\ell-1}) \sim N(0, \sigma_g^2(t_\ell - t_{\ell-1})).$$

Thus, we conclude that

$$W_n(t_j) \Rightarrow \sum_{\ell=1}^j \sum_{g=1}^S \sqrt{p_g} e^{\lambda_g(t_j - t_\ell)} e^{\lambda_g t_{\ell-1}/2} V_g(t_\ell - t_{\ell-1}) \text{ as } n \rightarrow \infty$$

Next we show that the random vector  $\mathbf{W}$  converges to the random vector  $\mathbf{Y}$ , which has the multivariate normal distribution. We can obtain the distribution for  $\mathbf{Y}$  from the independence between  $V_i(t_\ell - t_{\ell-1})$  and  $V_j(t_m - t_{m-1})$  for all  $i, j \in \{1, \dots, N_g\}, \ell, m \in \{1, \dots, k\}$ :

$$\mathbf{Y} = [Y(t_1), \dots, Y(t_k)] \sim N(0, \Sigma)$$

where

$$\Sigma_{i,j} = \sum_{\ell=1}^{\min(i,j)} \sum_{g=1}^S p_g e^{\lambda_g(t_i - t_\ell)} e^{\lambda_g(t_j - t_\ell)} e^{\lambda_g t_{\ell-1}} \sigma_k^2(t_\ell - t_{\ell-1})$$

Then we use the Cramer Wold device, given a constant vector  $a \in \mathbb{R}^k < \infty$ , we have

$$\begin{aligned} \langle a, \mathbf{W} \rangle &= \sum_{j=1}^k a_j \sum_{\ell=1}^j \sum_{g=1}^S \sqrt{p_g} e^{\lambda_g(t_j - t_\ell)} \left( \frac{X_g(t_{\ell-1})}{p_g n} \right)^{1/2} X_g(t_{\ell-1})^{-1/2} \sum_{m=1}^{X_g(t_{\ell-1})} \left[ B_m(t_\ell - t_{\ell-1}) - e^{\lambda_g(t_\ell - t_{\ell-1})} \right] \\ &\Rightarrow \sum_{j=1}^k a_j \sum_{\ell=1}^j \sum_{g=1}^S \sqrt{p_g} e^{\lambda_g(t_j - t_\ell)} e^{\lambda_g t_{\ell-1}/2} V_i(t_\ell - t_{\ell-1}) \\ &= \sum_{j=1}^k a_j Y(t_j) = \langle a, \mathbf{Y} \rangle \end{aligned}$$

Thus, we show that  $\mathbf{W} \Rightarrow \mathbf{Y}$ .

## 5.4 Proof of proposition 2 if initial proportions can go to 0 with $n$

In proving Proposition 2, we made the assumption that the initial proportions  $p_i$  for  $i = 1, \dots, S$  are not dependent on the initial cell count  $n$ . In this sub-section, we aim to relax this assumption and demonstrate a similar result. Note we will assume that all other inputs are still independent of  $n$ , i.e.,  $S$  and  $\mathcal{T}$ .

In particular, we will allow  $p_i$  to depend on  $n$ , and allow for  $\limsup_n p_i n < \infty$ . We denote two sets of subpopulations  $F$  and  $I$ , where

$$\begin{aligned} F &= \{i \in \{1, \dots, S\}; \limsup_{n \rightarrow \infty} p_i n < \infty\} \\ I &= \{i \in \{1, \dots, S\}; \lim_{n \rightarrow \infty} p_i n = \infty\} \end{aligned}$$

Due to  $\sum_{i=1}^S p_i = 1$ , and  $S$  being fixed with  $n$ , we know that the set  $I$  must not be an empty set. Then following a similar pattern as the proof of the Proposition 2, we derive:

$$\begin{aligned}
W_n(t_j) &= \sum_{\ell=1}^j \sum_{g=1}^S e^{\lambda_g(t_j-t_\ell)} \left(\frac{p_g n}{n}\right)^{1/2} \left(\frac{X_g(t_{\ell-1})}{p_g n}\right)^{1/2} X_g(t_{\ell-1})^{-1/2} \sum_{m=1}^{X_g(t_{\ell-1})} \left[B_g(t_\ell - t_{\ell-1}) - e^{\lambda_g(t_\ell-t_{\ell-1})}\right] \\
&= \sum_{\ell=1}^j \sum_{g \in I} e^{\lambda_g(t_j-t_\ell)} \left(\frac{p_g n}{n}\right)^{1/2} \left(\frac{X_g(t_{\ell-1})}{p_g n}\right)^{1/2} X_g(t_{\ell-1})^{-1/2} \sum_{m=1}^{X_g(t_{\ell-1})} \left[B_g(t_\ell - t_{\ell-1}) - e^{\lambda_g(t_\ell-t_{\ell-1})}\right] \\
&\quad + \sum_{\ell=1}^j \sum_{g \in F} e^{\lambda_g(t_j-t_\ell)} \left(\frac{p_g n}{n}\right)^{1/2} \left(\frac{X_g(t_{\ell-1})}{p_g n}\right)^{1/2} X_g(t_{\ell-1})^{-1/2} \sum_{m=1}^{X_g(t_{\ell-1})} \left[B_g(t_\ell - t_{\ell-1}) - e^{\lambda_g(t_\ell-t_{\ell-1})}\right]
\end{aligned}$$

Next, we may consider these two double sums separately. For  $g \in I$ , because the  $p_g n$  diverges to infinity as  $n \rightarrow \infty$ , the first double sum will converge to the same limit we established in Proposition 2. For  $g \in F$ ,  $p_g n$  will stay bounded. Therefore, in the second double sum, we will have  $\frac{p_g n}{n}$  converge to 0 as  $n \rightarrow \infty$ , which makes the second double sum vanish. In conclusion, we have

$$W_n(t_j) \Rightarrow \sum_{\ell=1}^j \sum_{g \in I} \sqrt{p_g} e^{\lambda_g(t_j-t_\ell)} e^{\lambda_g t_{\ell-1}/2} V_g(t_\ell - t_{\ell-1}) \text{ as } n \rightarrow \infty.$$

Note that this convergence result will lead to the same realization in practice, i.e., we will define the covariance by summing over all subtypes. This is because in practice we only have  $n < \infty$  and we cannot actually assume subpopulations have zero contribution to the covariance.

## 5.5 Exact path likelihood computation

Here we show how to calculate the exact likelihood of a sample path observation of the total cell count of multiple heterogeneous birth-death processes. While we do not use this approach for likelihood evaluation in the current manuscript we report it here to show that it is not feasible.

We first consider the following joint probability of a homogeneous linear birth-death process:

$$\mathbb{P}(X(t_1) = x_1, \dots, X(t_k) = x_k | X(t_0) = n, \theta_{BD}(2)) = \prod_{k=1}^{N_t} \mathbb{P}(X(t_k) = x_k | X(t_{k-1}) = x_{k-1}, \theta_{BD}(2)).$$

For ease of notation define the transition probability  $p_{i,j}(t_k - t_{k-1}) = \mathbb{P}(X(t_k) = j | X(t_{k-1}) = i, \theta_{BD}(2))$ . It is important to note that evaluating  $p_{i,j}(t_k - t_{k-1})$  is the most computation-demanding task when evaluating the joint probability. As a result, we mainly consider the number of evaluations of this transition probability. The analytical form for this transition probability was derived in [1]:

$$p_{i,j}(t) = \sum_{k=0}^{\min(i,j)} \binom{i}{k} \binom{i+j-k-1}{i-1} a(t)^{i-k} b(t)^{j-k} (1-a(t)-b(t))^k, \quad (15)$$

where

$$a(t) = \frac{\nu(e^{(\beta-\nu)t} - 1)}{\beta e^{(\beta-\nu)t} - \nu}, b(t) = \frac{\beta(e^{(\beta-\nu)t} - 1)}{\beta e^{(\beta-\nu)t} - \nu}.$$

Note that numerical evaluation (15) will be computationally expensive due to the presence of multiple factorial terms. We use a Gosper refined version of the Stirling formula [8] to approximate these

factorials

$$n! \approx \sqrt{\left(2n + \frac{1}{3}\right)} \pi n^n e^{-n}.$$

We find that this approximation leads to good performance in our examples. It is thus straightforward to evaluate the path likelihood for the case of a homogeneous linear birth-death process.

If we instead have observations of a sum of birth-death processes, the evaluation of the path likelihood is much more difficult. In particular, the sum of the birth-death processes is no longer a Markov process and we must therefore sum over possible values of our unobserved subpopulations. Specifically, if we have two subpopulations we can formulate the equation (10), as

$$\begin{aligned} P(X^{(r)}(t) = x_t, t \in \mathcal{T} | \theta_{BD}(2)) &= \sum_{i_1=0}^{x_1} \cdots \sum_{i_{N_T}=0}^{x_{N_T}} P(X_1(t_1) = i_1, X_2(t_1) = x_1 - i_1, \dots, X_1(t_{N_T}) = i_{N_T}, X_2(t_{N_T}) = x_{N_T} - i_{N_T} | \theta_{BD}(2)) \\ &= \sum_{i_1=0}^{x_1} \cdots \sum_{i_{N_T}=0}^{x_{N_T}} P(X_1(t_1) = i_1, X_1(t_2) = i_2, \dots, X_1(t_{N_T}) = i_{N_T} | \theta_{BD}(2)) \\ &\quad \times P(X_2(t_1) = x_1 - i_1, X_2(t_2) = x_2 - i_2, \dots, X_2(t_{N_T}) = x_{N_T} - i_{N_T} | \theta_{BD}(2)). \end{aligned} \tag{16}$$

Note that the last equality is due to the assumption that subpopulation grow independently, and we can decompose the joint probability of mixture cell count into a summation of multiple joint probabilities of the homogeneous linear birth-death process. It is not hard to see that if we naively evaluate the above sum, the number of computations of the homogeneous joint probability is around  $\Omega(\min_{t \in \mathcal{T}} x_t^{N_T})$ . Since many examples have  $N_T \approx 10$ , and  $x_t \approx 100$  this is clearly an infeasible approach.

In order to avoid the exponential dependence on the number of time points, one option is to use techniques from Hidden Markov Models (HMM). The main assumption of HMM is the Markov property of the hidden process, and that the hidden process relates to the observable process according to a specified distribution  $B$ . Recall that the time series of observed total cell count is given by  $\{X(t_i); i \in \mathcal{T}\}$  and denote the time series of the subpopulations as  $\{(X_1(t_i), \dots, X_S(t_i)); i \in \mathcal{T}\}$ . Then  $\{X(t_i); i \in \mathcal{T}\}$  in the HMM is the observable process, and  $\{(X_1(t_i), \dots, X_S(t_i)); i \in \mathcal{T}\}$  is the hidden Markov process due to the Markov property of the linear birth-death process. Notice that the relationship between the hidden process  $\{(X_1(t_i), \dots, X_S(t_i)); i \in \mathcal{T}\}$  and the observable process  $\{X(t_i); i \in \mathcal{T}\}$  can be defined as

$$\mathbb{P}(X(t) = x | (X_1(t), \dots, X_S(t)) = (x_1, \dots, x_S)) = \begin{cases} 1 & \text{if } x_1 + \dots + x_S = x \\ 0 & \text{o.w.} \end{cases}$$

We can translate the live-cell imaging experiment into an HMM and significantly improve the computational complexity of evaluating the exact likelihood function (7). In particular, we can use popular HMM techniques, such as the forward-backward procedure to reduce the total number of transition probability, i.e., equation (15), computation to  $\Theta(H^2 N_T)$  for one replicate at one dosage level, where  $H$  is the number of hidden states. In particular, we need to calculate the  $H$  by  $H$  transition matrix for every time point, and if we assume the length of time intervals are identical, we can reduce the upper bound to  $\Theta(H^2)$ . However, the number of hidden states depends on both the maximum total number of cells observed at each time point  $x$  and the number of subpopulations  $S$ .

As we will now show, this is unfortunately not a sufficient reduction in computational complexity. In particular, assume we observe  $x$  total cells. In that case, the number of hidden states is given by  $\binom{x-1}{S-1}$ , and assuming that  $x \gg S$ , we have that  $\binom{x-1}{S-1} = \Theta(x^{S-1})$  as  $x \rightarrow \infty$ . With only two subpopulations this results in computational complexity of  $\Theta(x^2)$ , and in many experiments, we might have  $x \approx 10^5$  leading to an extremely high computational burden. If  $S = 3$  we would end up with the far worse computational complexity of  $\Theta(x^4)$ .

In conclusion, using a naive approach to compute the joint probability of mixture cell count would require  $\Omega\left(\min_{t \in \mathcal{T}} x_t^{N_T}\right)$  many computations of transition probability, which is clearly infeasible when many cell counts are in the thousands with over 10 observed time points. We also discuss an HMM based approach to evaluating the likelihood that results in a significant reduction in the computational burden for evaluating this likelihood. In particular, with this approach we can reduce the number of computations of the transition probability to  $\Omega\left(\min_{t \in \mathcal{T}} x_t^2\right)$ . Unfortunately we might have  $\min_{t \in \mathcal{T}} x_t \approx 5000$ . Note that this will be the complexity for evaluating the likelihood of one single replicate at a single dose, so taking into account that we can have more than 10 different doses, with more than 10 replicates at each dose we see that unfortunately, this HMM approach is computationally infeasible. In addition, this HMM approach will have significantly worse computational complexity after we include observation noise terms and/or more than 2 subpopulations.

## 5.6 Data and code availability

All data and code used for running experiments, model fitting, and plotting is available on a GitHub repository at [https://github.com/chenyuwu233/PhenoPop\\_stochastic](https://github.com/chenyuwu233/PhenoPop_stochastic). The required Matlab version is Matlab R2022a or newer.

## Acknowledgements

The work of C. Wu was supported in part with funds from the Norwegian Centennial Chair Program. The work of K. Leder was supported in part with funds from NSF award CMMI 2228034 and Research Council of Norway Grant 309273. The work of EBG and JF was supported in part by NIH grant R01 CA241137, NSF DMS 2052465, NSF CMMI 2228034, and Research Council of Norway Grant 309273. The work of J.M. Enserink and D.S. Tadele was supported by grants from the Norwegian Health Authority South-East, grant numbers 2017064, 2018012, and 2019096; the Norwegian Cancer Society, grant numbers 182524 and 208012; and the Research Council of Norway through its Centers of Excellence funding scheme (262652) and through grants and 261936, 294916 and 314811

## References

- [1] Norman T.J. Bailey. *The elements of stochastic processes with applications to the natural sciences*, volume 25. John Wiley & Sons, 1991.
- [2] L. Baringhaus and C. Franz. On a new multivariate two-sample test. *Journal of Multivariate Analysis*, 88(1):190–206, 2004.
- [3] Cecilia Bonolo de Campos, Nathalie Meurice, Joachim L Petit, Alysia N Polito, Yuan Xiao Zhu, Panwen Wang, Laura A Bruins, Xuewei Wang, Isel D Lopez Armenta, Susie A Darvish,

- et al. “direct to drug” screening as a precision medicine tool in multiple myeloma. *Blood cancer journal*, 10(5):1–16, 2020.
- [4] Daniela Cadena-Herrera, Joshua E. Esparza-De Lara, Nancy D. Ramírez-Ibañez, Carlos A. López-Morales, Néstor O. Pérez, Luis F. Flores-Ortiz, and Emilio Medina-Rivero. Validation of three viable-cell counting methods: Manual, semi-automated, and automated. *Biotechnology Reports*, 7:9–16, 2015.
- [5] Harald Cramér. On the composition of elementary errors. *Scandinavian Actuarial Journal*, 1928(1):141–180, 1928.
- [6] Rick Durrett. *Branching process models of cancer*. Springer International Publishing, 2014.
- [7] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*. John Wiley & Sons, 2009.
- [8] R William Gosper Jr. Decision procedure for indefinite hypergeometric summation. *Proceedings of the National Academy of Sciences*, 75(1):40–42, 1978.
- [9] Carla Grandori and Christopher J Kemp. Personalized cancer models for target discovery and precision medicine. *Trends in cancer*, 4(9):634–642, 2018.
- [10] 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.
- [11] 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.
- [12] 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, October 2015 2016.
- [13] Rita Hannah, Michael Beck, Richard Moravec, and T Riss. Celltiter-glo? luminescent cell viability assay: a sensitive and rapid method for determining cell viability. *Promega Cell Notes*, 2:11–13, 2001.
- [14] M Kimmel and DE Axelrod. *Branching processes in biology*. Springer, New York, 2002.
- [15] Alvaro Köhn-Luque, Even Moa Myklebust, Dagim Shiferaw Tadele, Mariaserena Giliberto, Leonard Schmiester, Jasmine Noory, Elise Harivel, Polina Arsenteva, Shannon M. Mumenthaler, Fredrik Schjervold, Kjetil Taskén, Jorrit M. Enserink, Kevin Leder, Arnaldo Frigessi, and Jasmine Foo. Phenotypic deconvolution in heterogeneous cancer cell populations using drug-screening data. *Cell Reports Methods*, page 100417, 2023.
- [16] Andriy Marusyk, Vanessa Almendro, and Kornelia Polyak. Intra-tumour heterogeneity: a looking glass for cancer? *Nature reviews cancer*, 12(5):323–334, 2012.
- [17] Matlab optimization toolbox, R2022a. The MathWorks, Natick, MA, USA.

- [18] Shannon M Matulis, Vikas A Gupta, Paola Neri, Nizar J Bahlis, Paulo Maciag, Joel D Levenson, Leonard T Heffner, Sagar Lonial, Ajay K Nooka, Jonathan L Kaufman, et al. Functional profiling of venetoclax sensitivity can predict clinical response in multiple myeloma. *Leukemia*, 33(5):1291–1296, 2019.
- [19] Shannon M Mumenthaler, Jasmine Foo, Kevin Leder, Nathan C Choi, David B Agus, William Pao, Parag Mallick, and Franziska Michor. Evolutionary modeling of combination treatment strategies to overcome resistance to tyrosine kinase inhibitors in non-small cell lung cancer. *Molecular pharmaceuticals*, 8(6):2069–2079, 2011.
- [20] Anthony G Pakes. Ch. 18. biological applications of branching processes. *Handbook of statistics*, 21:693–773, 2003.
- [21] Chantal Pauli, Benjamin D Hopkins, Davide Prandi, Reid Shaw, Tarcisio Fedrizzi, Andrea Sboner, Verena Sailer, Michael Augello, Loredana Puca, Rachele Rosati, et al. Personalized in vitro and in vivo cancer models to guide precision medicine. *Cancer discovery*, 7(5):462–477, 2017.
- [22] Tea Pemovska, Mika Kontro, Bhagwan Yadav, Henrik Edgren, Samuli Eldfors, Agnieszka Szwa-jda, Henrikki Almusa, Maxim M Beshpalov, Pekka Ellonen, Erkki Elonen, et al. Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemiaism approach to therapy selection. *Cancer discovery*, 3(12):1416–1429, 2013.
- [23] Nikita Pozdeyev, Minjae Yoo, Ryan Mackie, Rebecca E Schweppe, Aik Choon Tan, and Bryan R Haugen. Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies. *Oncotarget*, 7(32):51619, 2016.
- [24] Gábor J Székely. E-statistics: The energy of statistical samples. *Bowling Green State University, Department of Mathematics and Statistics Technical Report*, 3(05):1–18, 2003.

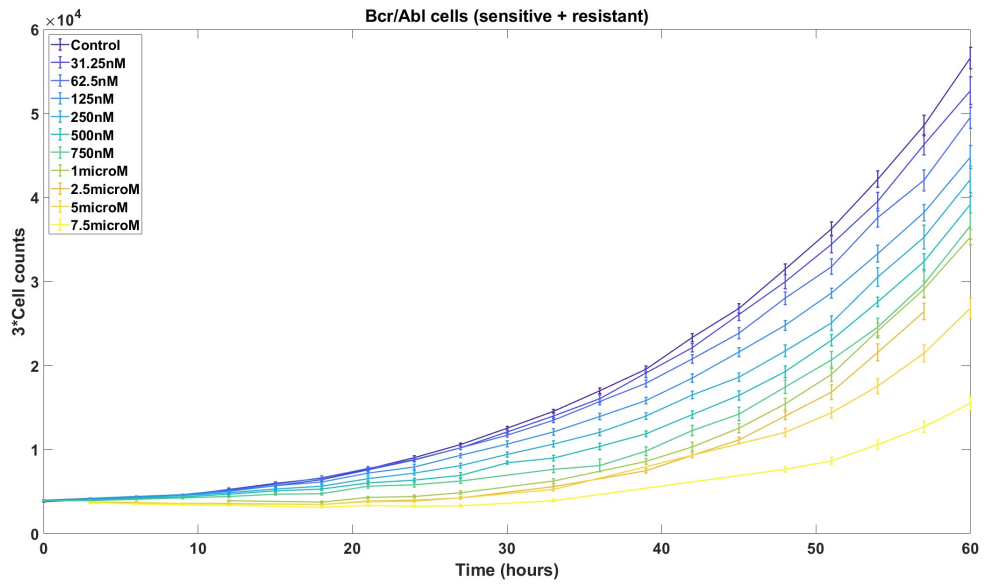


Figure 1: One to one mixtures of imatinib-sensitive and resistant Ba/F3 cells are counted at 14 different time points under 11 different concentrations of imatinib. Error bars, based on 14 replicates with outliers removed, depict the sample standard deviation, which increase with larger cell counts.

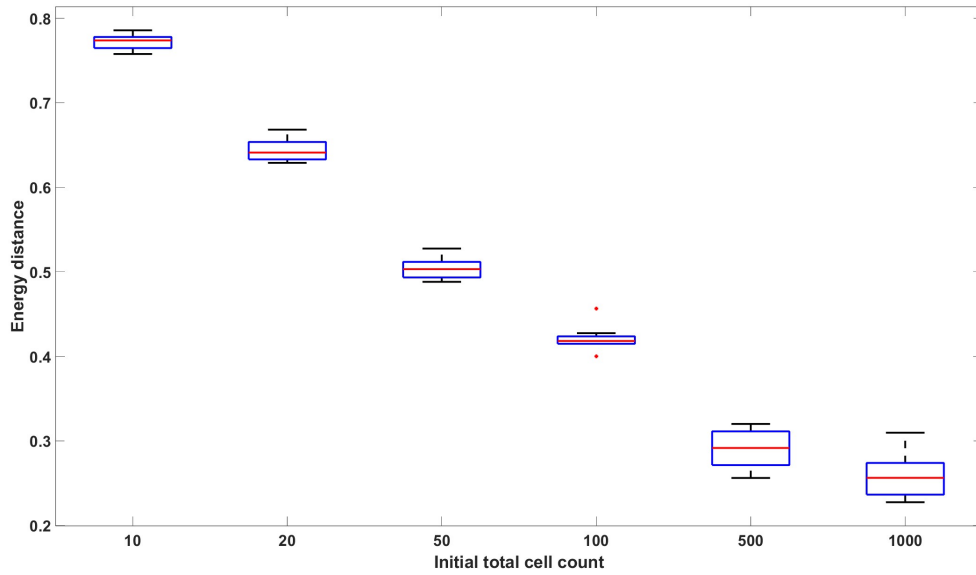


Figure 2: Empirical energy distance between linear birth-death simulated data and multivariate normal distributed data with respect to varying initial cell count:  $[10, 20, 50, 100, 500, 1000]$ . The data consists of  $N_R = 100,000$  replicates and 7 time points  $\mathcal{T} = [1, 2, 3, 4, 5, 6, 7]$ . No drug effect is assumed. The parameters used to generate the data are  $p_1 = 0.4629, \beta_1 = 0.9058, \nu_1 = 0.8101, p_2 = 0.5371, \beta_2 = 0.2785, \nu_2 = 0.2300$ . The box plot represents the values from 10 distinct datasets. The figure demonstrates that the distribution of the linear birth-death process converges to the multivariate normal distribution with mean and covariance given by Proposition 2 as the initial cell count increases.



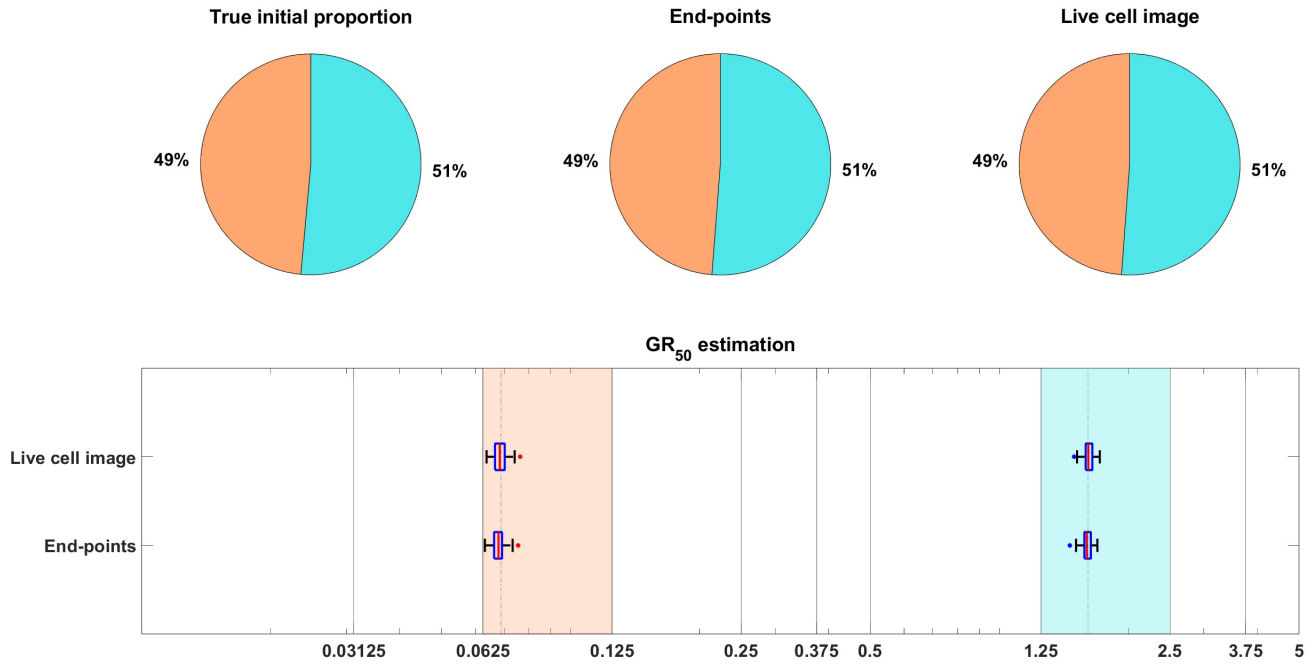


Figure 3: Estimation of the initial proportion and  $GR_{50}$  for 2 subpopulations using the end-points method and the live cell image method on simulated data. The parameter vector  $\theta_{BD}(2)$  and observation noise  $c$  used in this example are  $p_s = 0.4856, \beta_s = 0.1163, \nu_s = 0.0176, b_s = 0.8262, E_s = 0.0674, m_s = 4.5404, p_r = 0.5144, \beta_r = 0.4624, \nu_r = 0.3978, b_r = 0.8062, E_r = 1.5776, m_r = 4.2002, c = 1.2103$ . The pie chart illustrates the average of all bootstrap estimates for the initial proportion, while the box plot summarizes the distribution of the estimates for the  $GR_{50}$ 's. The vertical dashed lines in the box plot correspond to the true  $GR_{50}$  values employed to generate the data, while the vertical solid lines indicate the concentration levels at which the data were collected. Each color in the plot represents a distinct subpopulation: orange for sensitive and blue for resistant. The shaded areas in the box plot indicate the concentration intervals where the true  $GR_{50}$ 's are located, and the colored dots mark outliers in the estimation of the  $GR_{50}$  for each subpopulation, with red for sensitive and blue for resistant. This example demonstrates that our newly proposed models can accurately recover the initial proportion and  $GR_{50}$  values with high precision.

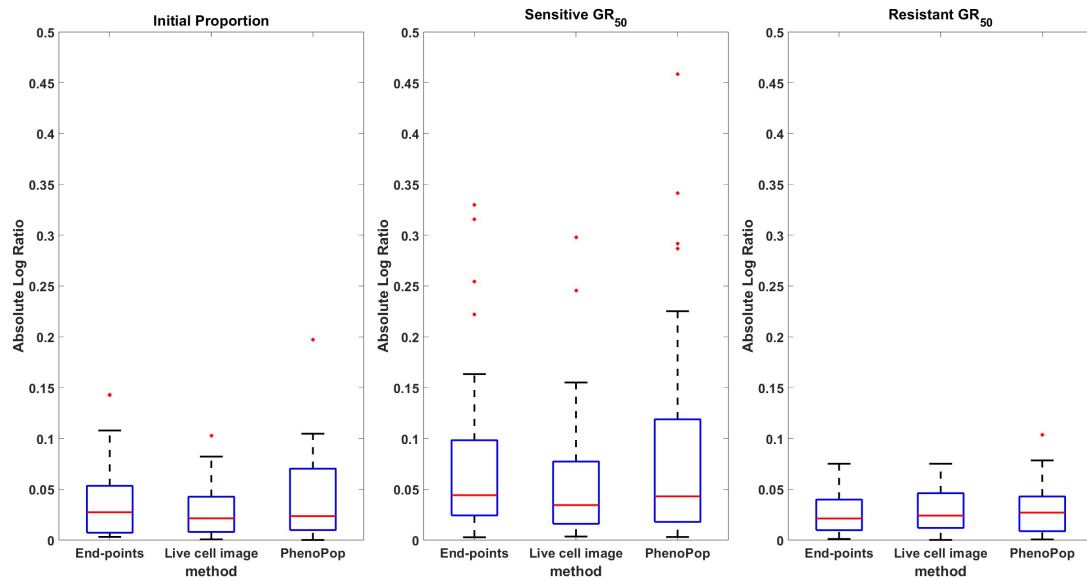


Figure 4: Absolute log ratio accuracy of three estimators  $\{\hat{p}_s, \hat{G}R_s, \hat{G}R_r\}$  using the PhenoPop, end-points and live cell image methods. The results are summarized based on 30 different simulated datasets. This figure demonstrates that there are no significant differences in estimation accuracy among these three methods when the true parameters fall within the range described in Table 1.

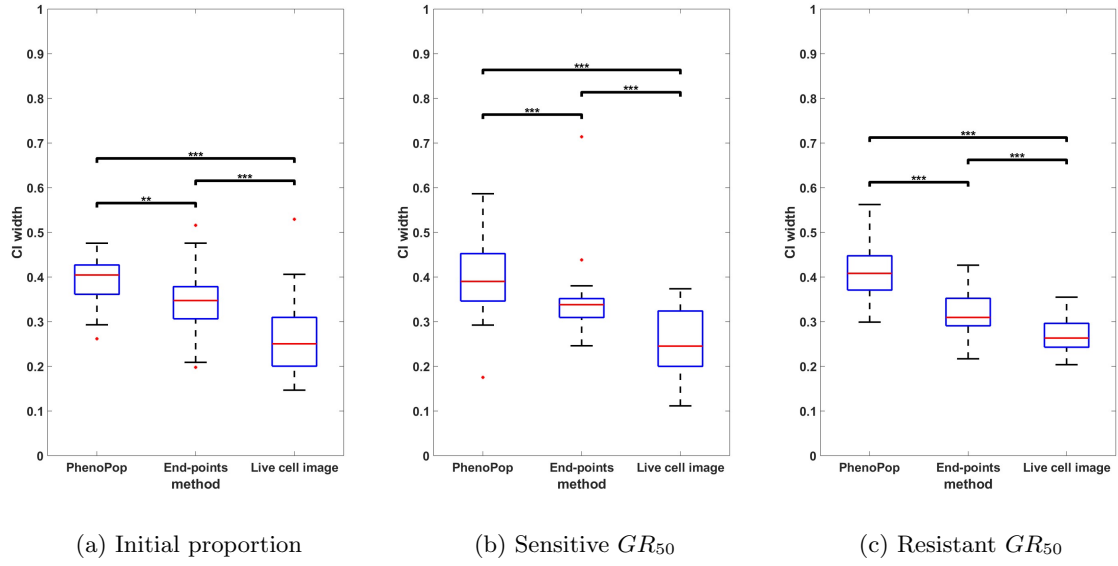


Figure 5: Comparison of the normalized CI widths of three estimators  $\{\hat{p}_s, \hat{G}R_s, \hat{G}R_r\}$  estimated from three different methods. The  $y$ -axis represents the normalized CI width. The box plot summarizes the results across 30 different simulated datasets. The significance bar indicates the p-values derived from the Wilcoxon rank-sum test, with significance levels denoted as  $*** \leq 0.001 \leq ** \leq 0.01 \leq * \leq 0.05$ . This figure demonstrates that the newly proposed models exhibit significant advantages in estimation precision, with the live cell image method demonstrating the highest level of precision.

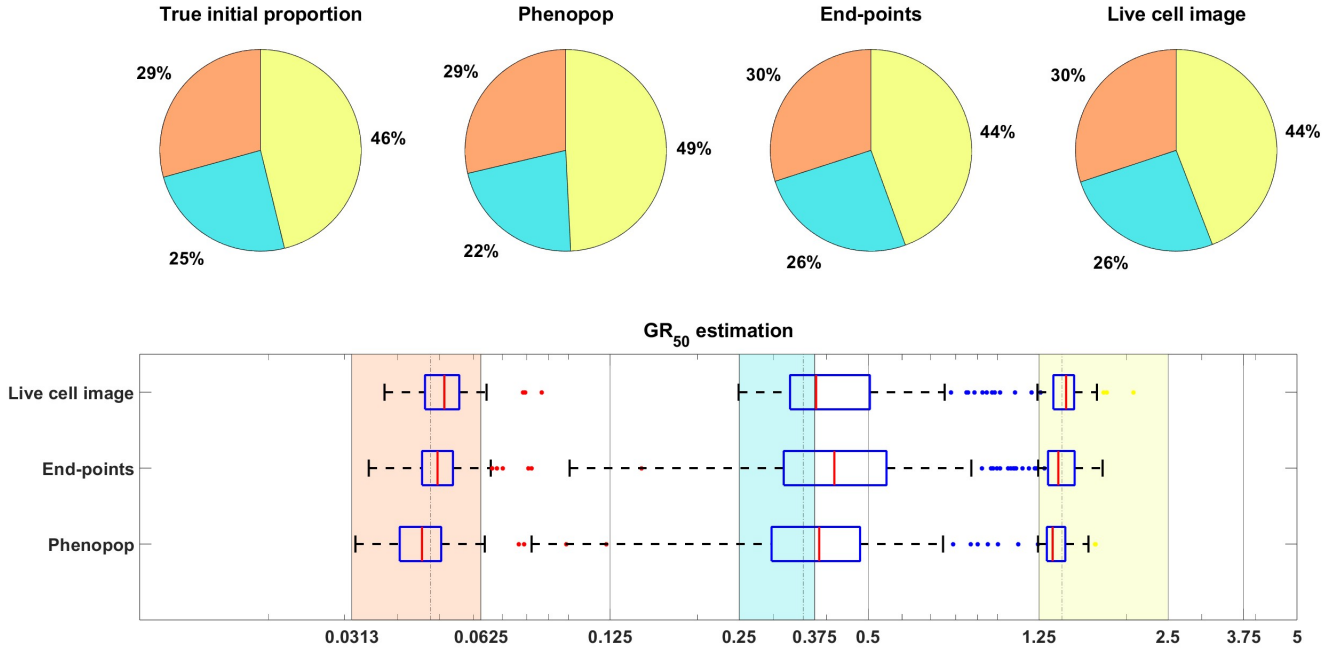


Figure 6: Estimation of the initial proportion and  $GR_{50}$  for 3 subpopulations using the three estimation methods. The parameter vector  $\theta_{BD}(3)$  and the observation noise in this example are  $p_s = 0.2135, \beta_s = 0.3214, \nu_s = 0.2773, b_s = 0.8782, E_s = 0.0344, m_s = 2.5998, p_m = 0.2718, \beta_m = 0.7334, \nu_m = 0.6776, b_m = 0.8506, E_m = 0.3558, m_m = 4.6055, p_r = 0.5147, \beta_r = 0.0683, \nu_r = 0.0253, b_r = 0.8614, E_r = 1.5764, m_r = 4.4706, c = 9.5209$ . The pie chart illustrates the average of all bootstrap estimates for the initial proportion, while the box plot summarizes the distribution of all estimates for the  $GR_{50}$ 's. The vertical dashed lines in the box plot correspond to the true  $GR_{50}$  values employed to generate the data, while the vertical solid lines indicate the concentration levels at which the data were collected. Each color in the plot represents a distinct subpopulation: orange for sensitive, blue for moderate, and yellow for resistant. The shaded areas in the box plot indicate the concentration intervals where the true  $GR_{50}$ 's are located, and the colored dots mark outliers in the estimation of the  $GR_{50}$  for each subpopulation, with red for sensitive, blue for moderate, and yellow for resistant.

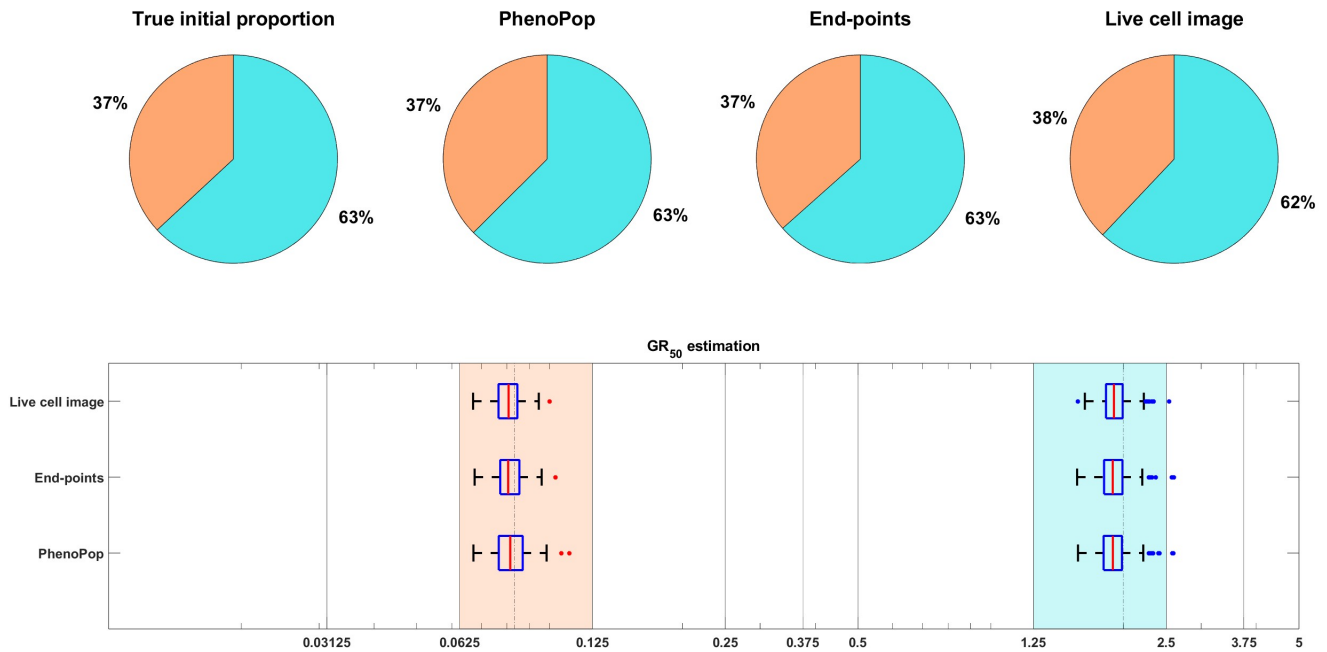


Figure 7: An illustrative example under the high observation noise scenario, i.e.  $c = 500$ . The parameter vector  $\theta_{BD}(2)$  and the observation noise in this example are  $p_s = 0.3690, \beta_s = 0.4380, \nu_s = 0.3422, b_s = 0.8398, E_s = 0.0813, m_s = 3.9647, p_r = 0.6310, \beta_r = 0.5320, \nu_r = 0.4767, b_r = 0.8674, E_r = 1.9793, m_r = 4.8357, c = 500$ . Results are presented as in Figure 3. This example demonstrates that all three methods are capable of recovering the initial proportion and  $GR_{50}$  even under the high observation noise scenario.

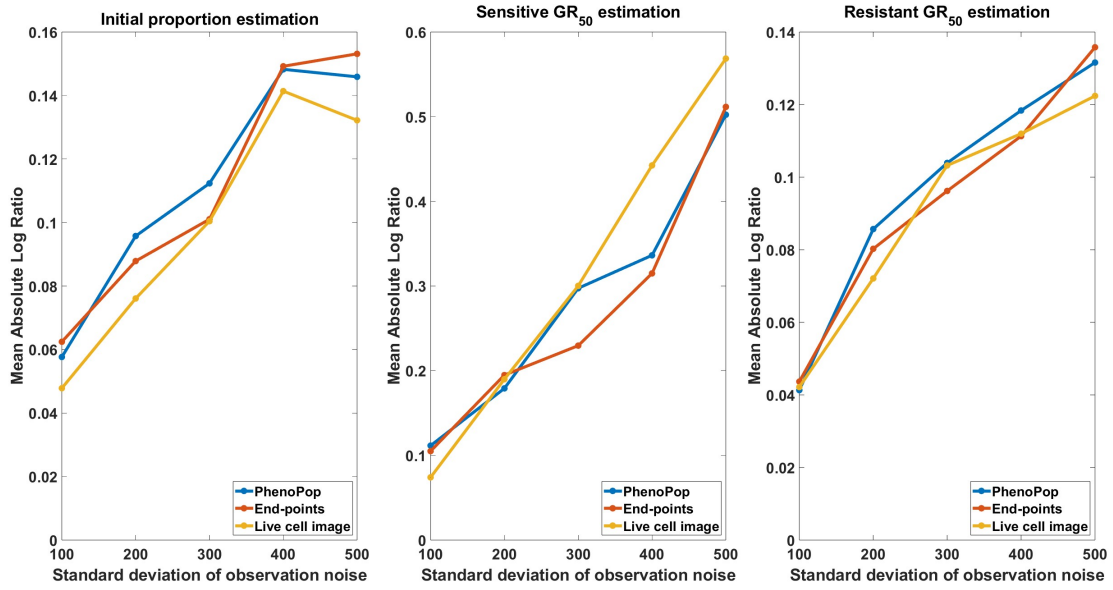


Figure 8: Estimation error of  $\{\hat{p}_s, \hat{G}R_s, \hat{G}R_r\}$  with respect to varying standard deviation of observation noise. The metric of estimation error is the mean absolute log ratio of estimates across 30 simulated datasets, each generated from a distinct parameter set. The value of the observation noise parameter,  $c$ , in these 30 generating parameter sets was assigned to 5 different values in the set  $\mathcal{C} = \{100, 200, 300, 400, 500\}$  to generate the line plot. Three different line plots correspond to three different methods, as indicated by the figure legends. This figure demonstrates that the estimations of the three methods deteriorate as the level of observation noise increases.

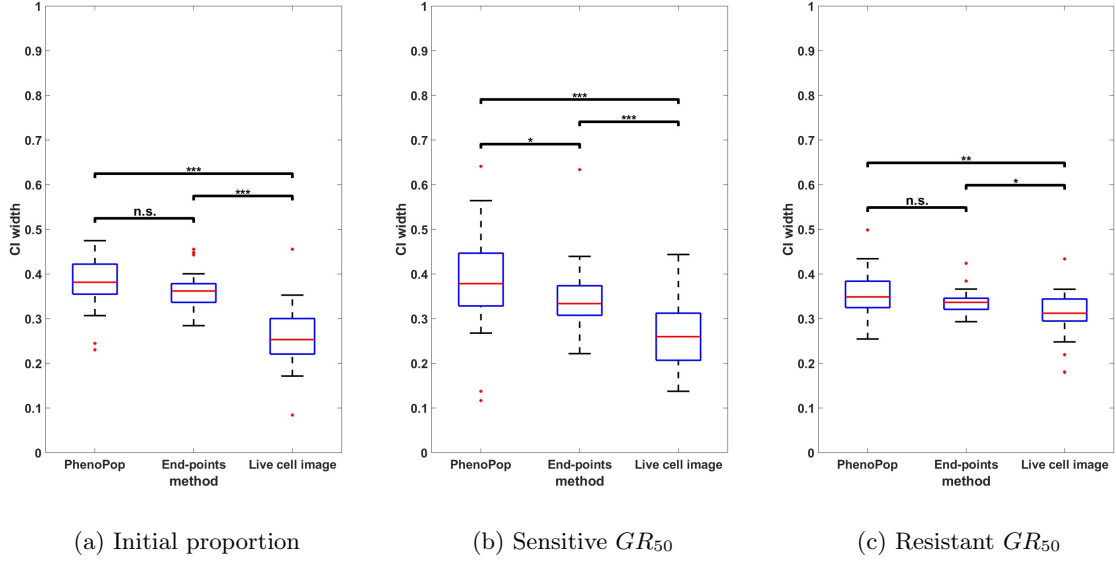


Figure 9: Comparison of the normalized CI widths of the three estimators  $\{\hat{p}_s, \hat{G}R_s, \hat{G}R_r\}$  using the three different estimation methods, when the observation noise parameter is set to  $c = 100$ . The  $y$ -axis represents the normalized CI width. The box plot summarizes the results across 30 different datasets. The significance bar indicates the p-values derived from the Wilcoxon rank-sum test, with significance levels denoted as  $*** \leq 0.001 \leq ** \leq 0.01 \leq * \leq 0.05$ . This figure demonstrates the advantages of the live cell image method in estimation precision are preserved even when the standard deviation of observation noise is 10% of the initial cell count.

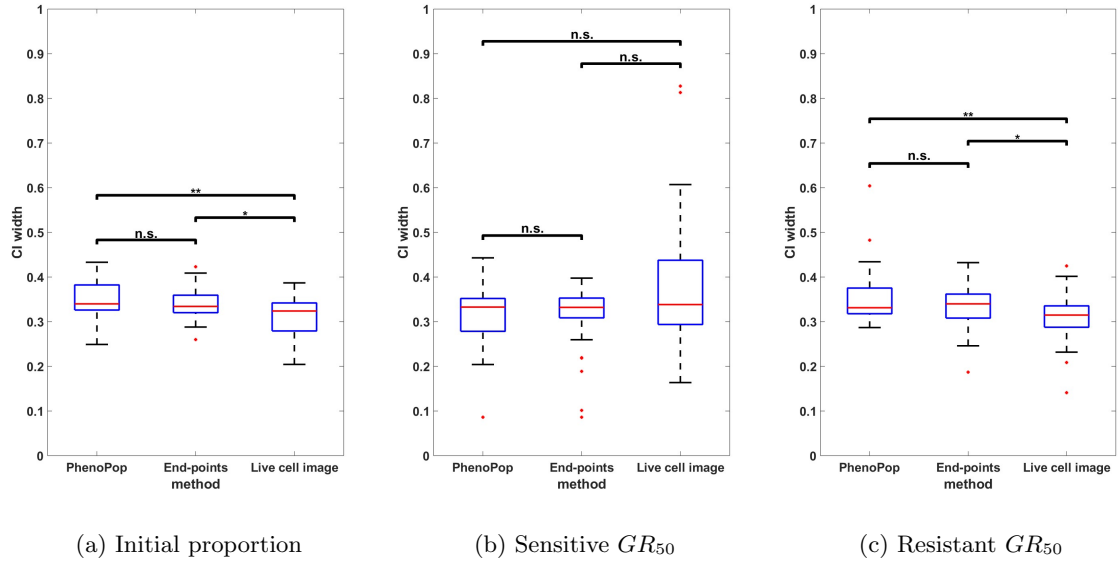


Figure 10: Comparison of the normalized CI widths of three estimators  $\{\hat{p}_s, \hat{G}R_s, \hat{G}R_r\}$  using the three different estimation methods, when the observation noise parameter is set to  $c = 500$ . Results are presented as in Figure 9. This figure demonstrates that the advantages of the live cell image method in estimation precision become less significant as the standard deviation of observation noise increases to 50% of the initial cell count.



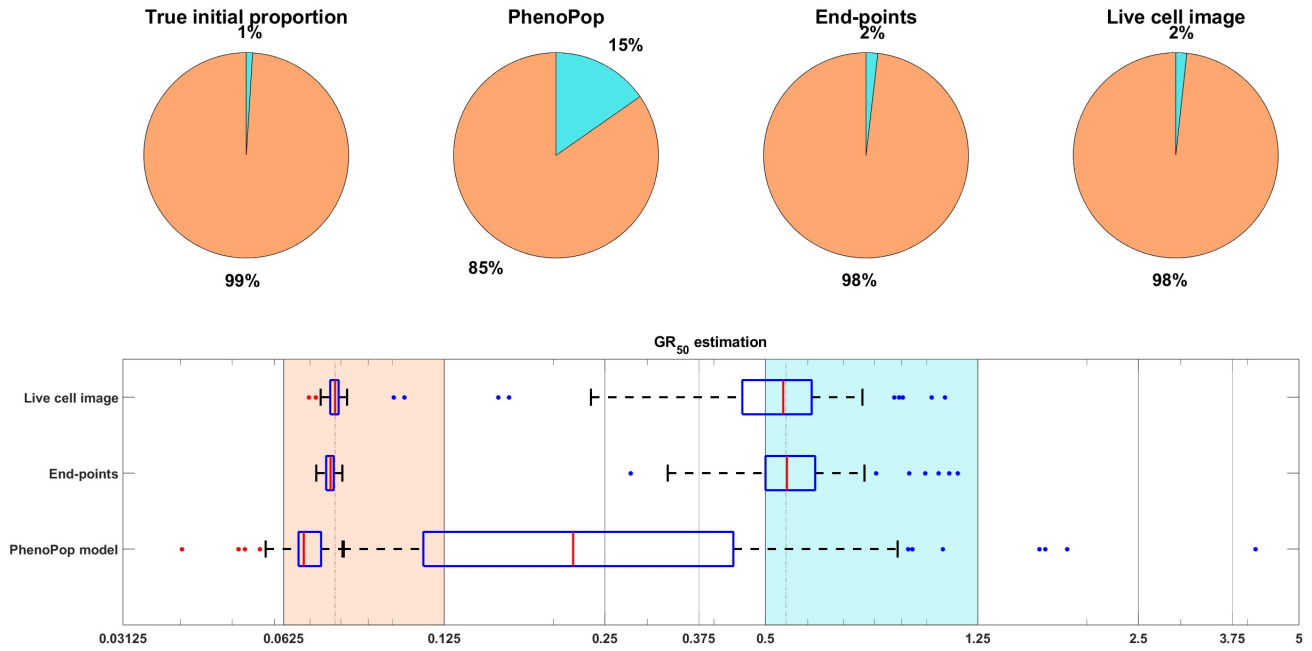


Figure 11: An illustrative example under the unbalanced initial proportion scenario, i.e.  $p_s = 0.99$ . The parameter vector  $\theta_{BD}(2)$  and the observation noise in this example are  $p_s = 0.9900, \beta_s = 0.4301, \nu_s = 0.4199, b_s = 0.8644, E_s = 0.0768, m_s = 4.3186, p_r = 0.0100, \beta_r = 0.1458, \nu_r = 0.1258, b_r = 0.8565, E_r = 0.5348, m_r = 3.7518, c = 4.8400$ . Results are presented as in Figure 3. This example demonstrates that our newly proposed model can accurately estimate parameters even when the initial proportion of the resistant subpopulation is negligible, while the PhenoPop method fails to estimate the parameters accurately.

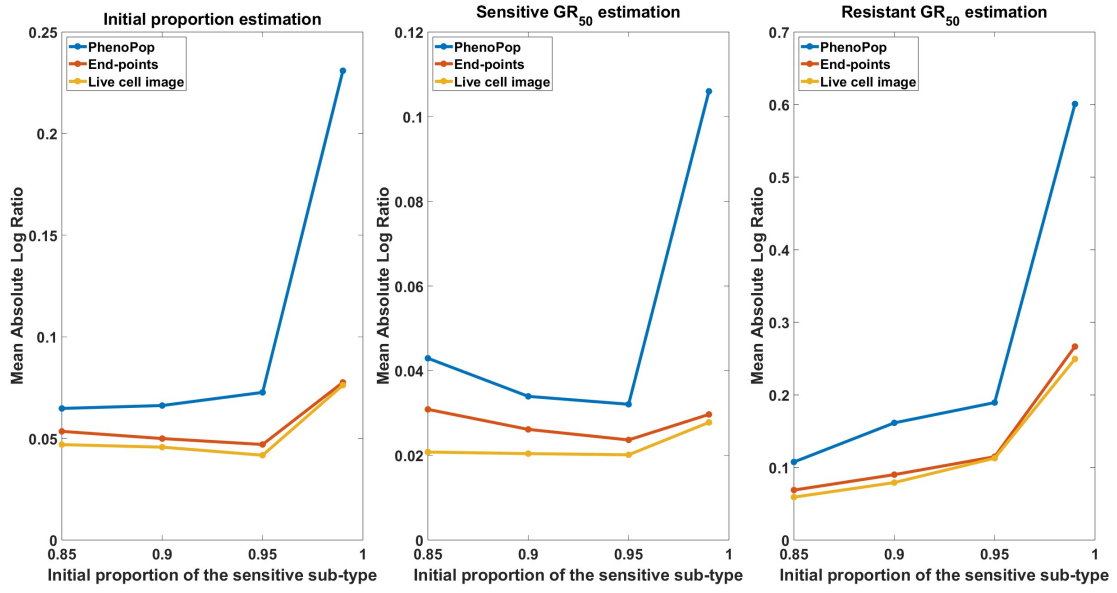


Figure 12: Estimation error of  $\{\hat{p}_s, \hat{G}R_s, \hat{G}R_r\}$  with respect to varying resistant initial proportions. The metric of estimation error is the mean absolute log ratio across 100 simulated datasets, each generated from a distinct parameter set. The value of  $p_s$  in these 100 generating parameter sets was assigned to 4 different values in the set  $\mathcal{P} = \{0.85, 0.90, 0.95, 0.99\}$  to generate the line plot. Three different line plots correspond to three different methods, as indicated by the figure legends. This figure demonstrates the advantages of estimation accuracy provided by the newly proposed methods when the initial proportion of the resistant subpopulation decreases toward 0.

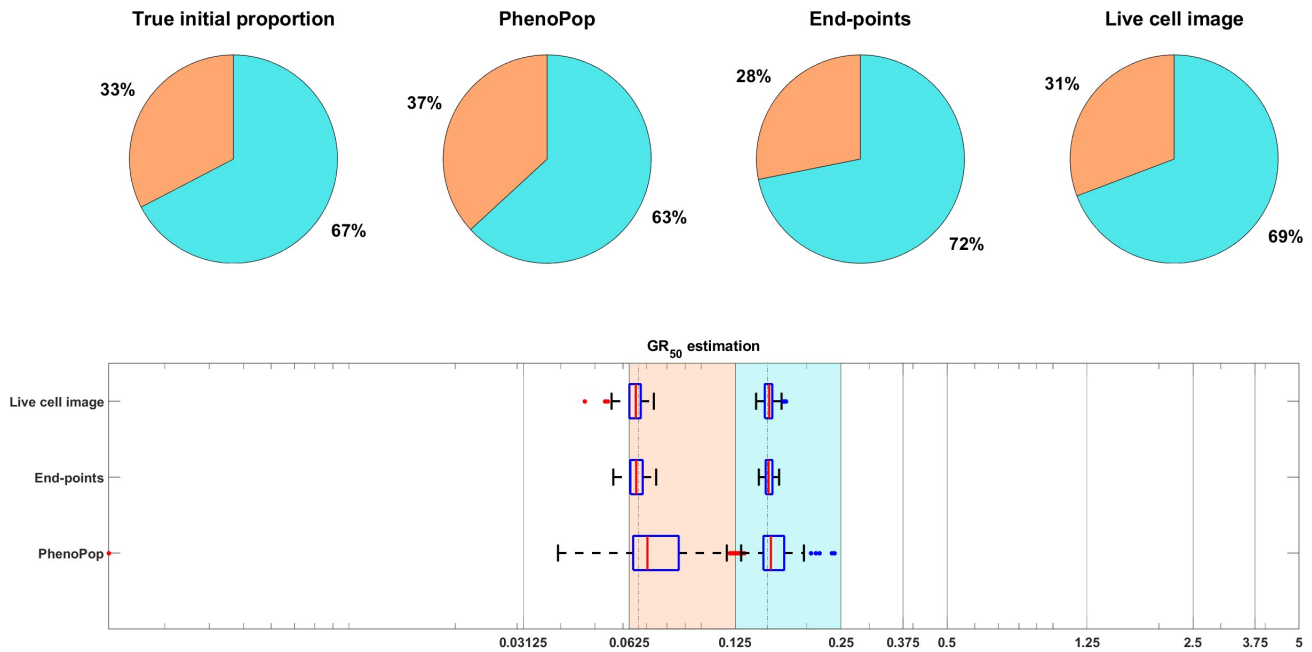


Figure 13: An illustrative example under the similar subpopulation sensitivity scenario. The parameter vector  $\theta_{BD}(2)$  and the observation noise in this example are  $p_s = 0.3263, \beta_s = 0.8896, \nu_s = 0.8215, b_s = 0.8820, E_s = 0.0654, m_s = 3.8539, p_r = 0.6737, \beta_r = 0.0925, \nu_r = 0.0661, b_r = 0.8171, E_r = 0.1500, m_r = 3.6015, c = 7.6660$ . Results are presented as in Figure 3. This example demonstrates that all three methods are capable of recovering the initial proportion and  $GR_{50}$  even when two subpopulations have similar drug sensitivity, while the newly proposed methods exhibit superior estimation precision compared to the PhenoPop method.

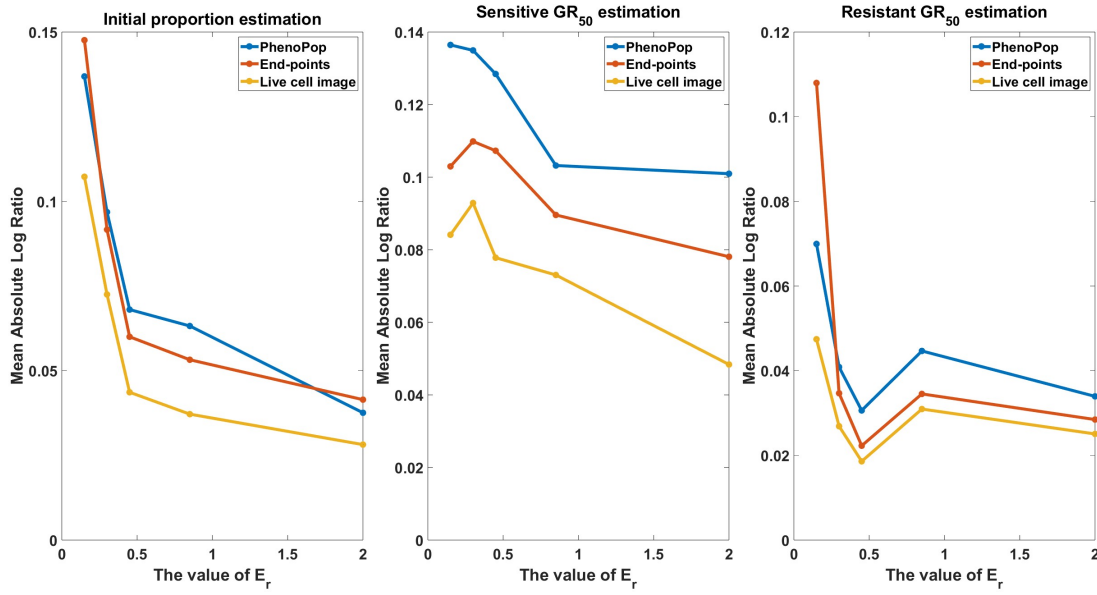


Figure 14: Estimation error of  $\{\hat{p}_s, \hat{G}R_s, \hat{G}R_r\}$  with respect to varying similarity between subpopulation drug sensitivities. The metric of estimation error is the mean absolute log ratio across 80 simulated datasets, each generated from a distinct parameter set. The value of  $E_r$  in these 80 generating parameter sets was assigned to 5 different values in the set  $\mathcal{E} = \{0.15, 0.3, 0.45, 0.85, 2\}$  to generate the line plot. Three different line plots correspond to three different methods, as indicated by the figure legends. This figure demonstrates that the estimation accuracy of the three methods improves as the discrepancy of drug sensitivity between the two subpopulations increases, with the live cell image method exhibiting the smallest average error among the three methods.