PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

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

Chenyu Wu[1], Einar Bjarki Gunnarsson[2,3], Even Moa Myklebust[4], Alvaro Köhn-Luque[4,5], Dagim Shiferaw Tadele[6,7,8], Jorrit Martijn Enserink[6,7,9], Arnoldo Frigessi[4], Jasmine Foo[2], Kevin Leder[1]*

1 Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, Minnesota, United States of America, **2** School of Mathematics, University of Minnesota, Minneapolis, Minnesota, United States of America, **3** Applied Mathematics Division, Science Institute, University of Iceland, Reykjavík, Iceland, **4** Oslo Centre for Biostatistics and Epidemiology, Faculty of Medicine, University of Oslo, Oslo, Norway, **5** Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway, **6** Department of Molecular Cell Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway, **7** Centre for Cancer Cell Reprogramming, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway, **8** Department of Medical Genetics, Oslo University Hospital, Oslo, Norway, **9** Section for Biochemistry and Molecular Biology, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway

* lede0024@umn.edu

Citation: Wu C, Gunnarsson EB, Myklebust EM, Köhn-Luque A, Tadele DS, Enserink JM, et al. (2024) Using birth-death processes to infer tumor subpopulation structure from live-cell imaging drug screening data. PLoS Comput Biol 20(3): e1011888. https://doi.org/10.1371/journal.pcbi.1011888

Editor: David Basanta, H. Lee Moffitt Cancer Center & Research Institute, UNITED STATES

Received: June 14, 2023

Accepted: February 4, 2024

Published: March 6, 2024

Data Availability Statement: All data and code used for running experiments, model fitting, and plotting are available on a GitHub repository at https://github.com/chenyuwu233/PhenoPop_stochastic. The required Matlab version is Matlab R2022a or newer. Stimulated data available at https://github.com/chenyuwu233/PhenoPop_stochastic/tree/main/In%20silico%20experiment In vitro data available at https://github.com/chenyuwu233/PhenoPop_stochastic/tree/main/In%20vitro%20experiment.

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.

Author summary

One of the main reasons tumors can be difficult to treat is the presence of multiple subpopulations each with a distinct response to a given therapy. In particular some of these subpopulations are able to evade anti-cancer therapies and give rise to treatment resistant

Funding: The work of C.W. was supported in part with funds from the Norwegian Centennial Chair Program and NSF award CMMI 2228034. The work of K.L. was supported in part with funds from NSF award CMMI 2228034 and Research Council of Norway Grant 309273. The work of E.B.G. Gunnarsson and J.F. 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.E. and D.S. T. 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. None of the sponsors or funders played any role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

disease. Therefore it is vitally important to be able to identify these subpopulations and furthermore quantify their response to a therapy of interest, i.e., to quantify a tumors subpopulation structure. A potential tool for quantifying a tumors subpopulation structure are so called high-throughput drug screens (HTDS). In these screens a patients tumor sample is collected and then conditioned to grow *in vitro* where it can be exposed to a variety of drugs at different concentration levels. In the present work we develop statistical tools that create quantitative estimates of tumor population substructure based on HTDS data. These estimators have better precision than previous results, and furthermore are able to more accurately identify smaller subpopulations than previous estimators.

Introduction

In recent years the design of personalized anti-cancer therapies has been greatly aided by the use of high throughput drug screens (HTDS) [1, 2]. In these studies a large panel of drugs is tested against a patient's tumor sample to identify the most effective treatment [3–6]. 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 [7]. 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 [8]. 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 [9–11], 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

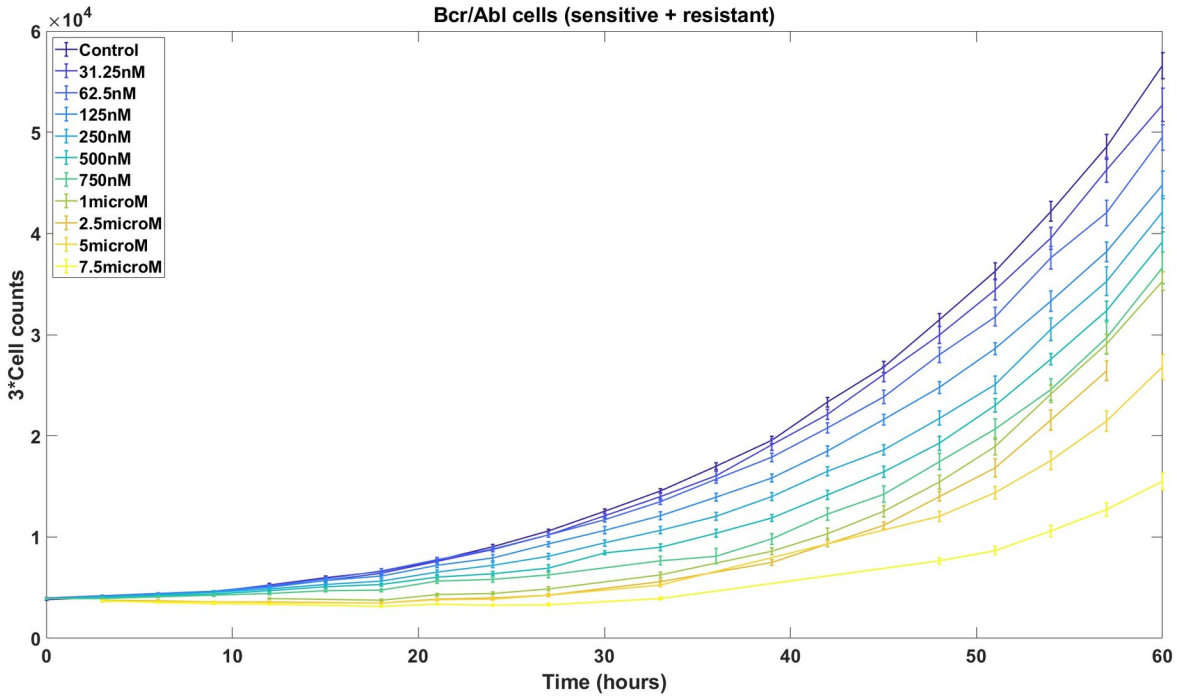

Fig 1. One to one mixtures of imatinib-sensitive and resistant Ba/F3 cells are counted at 14 different time points under 11 different concentrations of imatinib. Error bars, based on 14 replicates with outliers removed, depict the sample standard deviation, which increases with larger cell counts.

https://doi.org/10.1371/journal.pcbi.1011888.g001

population size (see Fig 1 and discussion in S1 Text). 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 Material and methods, 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 Result, 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 Conclusion.

Material and methods

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 $\mathcal{T} = \{t_1, \ldots, t_{N_T}\}$ and drug concentrations $\mathcal{D} = \{d_1, \ldots, d_{N_D}\}$. 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

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

PhenoPop for drug response deconvolution in cell populations

In [8], 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

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

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

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

Note that $H(0; b, E, m) = 1$ and $H(d; b, E, m) \to b$ as $d \to \infty$. Therefore, the population grows at exponential rate α in the absence of drug ($d = 0$) and at rate $\alpha + \log(b) < \alpha$ for an arbitrarily large drug dose ($d \to \infty$). In other words, $\log(b)$ represents the theoretical maximum drug effect on the growth rate. The parameter E represents the dose at which the drug has half the theoretical 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 α_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

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

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

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

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

$$\sigma^2_{hl}(t, d) = \begin{cases} \sigma^2_H, & t \geq T_L \text{ and } d \leq D_L \\ \sigma^2_L, & \text{otherwise.} \end{cases}$$

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 [8]. 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 $\{Z^{(r)}(t,d); r \in \{1,\ldots,N_R\}\}$ are independent random variables with the normal distribution $N(0, \sigma^2_{hl}(t,d))$. This model has the parameter set

$$\theta_{PP}(S) = \{(p_i, \alpha_i, b_i, E_i, m_i), \sigma_H, \sigma_L; i \in \{1,\ldots,S\}\}. \quad (2)$$

The initial fractions of the S subpopulations $\{p_i : i \in \{1,\ldots S\}\}$ and the parameters $\{(\alpha_i, b_i, E_i, m_i) : i \in \{1,\ldots,S\}\}$ governing the drug responses of the subpopulations are unknown. In addition, the variance levels σ^2_H and σ^2_L 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

$$L_{PP}(\theta_{PP}(S)|\mathbf{x}) = \prod_{r=1}^{N_R} \prod_{(t,d)\in\mathcal{T}\times\mathcal{D}} \frac{1}{\sqrt{2\pi\sigma^2_{hl}(t,d)}} \exp\left[-\frac{(x_{t,d,r} - f(t,d))^2}{2\sigma^2_{hl}(t,d)}\right]. \quad (3)$$

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 [8].

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 Fig 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.

Linear birth-death process

A natural extension of PhenoPop [8] 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

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

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 α_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 [12]

$$\mathbb{E}[X_i(t,d)] \doteq np_i\mu_i(t,d) = np_i e^{\lambda_i(d)t} \tag{4}$$

$$\mathrm{Var}[X_i(t,d)] \doteq np_i\sigma_i^2(t,d) = np_i \frac{\beta_i + \nu_i(d)}{\lambda_i(d)} \left(e^{2\lambda_i(d)t} - e^{\lambda_i(d)t}\right). \tag{5}$$

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

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

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

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

$$x_{t,d,r} = X^{(r)}(t,d) + Z_{t,d,r}, \tag{6}$$

where $X^{(r)}(t, d)$ are independent copies of $X(t, d)$ for $r = 1, \ldots, N_R$, and $\{Z_{t,d,r}; d \in \mathcal{D}, t \in \mathcal{T}, r \in \{1, \ldots, N_R\}\}$ 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

$$\theta_{BD}(S) = \{(p_i, \beta_i, \nu_i, b_i, E_i, m_i), c; i \in \{1, \ldots, S\}\}. \tag{7}$$

In comparison with PhenoPop, on the one hand, the growth rate parameter α_i for each subpopulation has been replaced by the birth and death rates β_i and ν_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

$$\begin{aligned} &L_{BD}(\theta_{BD}(S)|\mathbf{x}) \\ &= \prod_{r=1}^{N_R} \prod_{d\in\mathcal{D}} P\Big(X^{(r)}(t,d) + Z_{t,d,r} \in (x_{t,d,r} - \Delta x_{t,d,r}, x_{t,d,r} + \Delta x_{t,d,r}), t \in \mathcal{T} | \theta_{BD}(S)\Big) \end{aligned} \tag{8}$$

where we assume that observations at different doses and from distinct replicates are independent, and $(x_{t,d,r} - \Delta x_{t,d,r}, x_{t,d,r} + \Delta x_{t,d,r})$ 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.

End-point experiments. For many common cell counting techniques, e.g. CellTiter-Glo [13], the experiment must be stopped to perform the viability assay and the cells are then killed. In this case, observations at different time points are actually observations of independent cell populations. We can therefore assume that conditional on the parameter $\theta_{BD}(S)$ these observations are independent. Thus, the likelihood function can be written as

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

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

$$W_n(t,d) = \frac{1}{\sqrt{n}} \sum_{i=1}^{S} \Big(X_i(t,d) - np_i e^{\lambda_i(d)t}\Big). \tag{9}$$

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 $n \to \infty$.*

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

$$L_{EP}(\theta_{BD}(S)|\mathbf{x}) = \prod_{r=1}^{N_R} \prod_{(t,d)\in\mathcal{T}\times\mathcal{D}} \frac{1}{\sqrt{2\pi(n\sigma^2(t,d) + c^2)}} \exp\left[-\frac{(x_{t,d,r} - \mu(t,d))^2}{2(n\sigma^2(t,d) + c^2)}\right]. \tag{10}$$

The right-hand side of (10) depends on the model parameters $\theta_{BD}(S)$ via the mean and variance functions $\mu(t, d)$ and $\sigma^2(t, d)$. As in [8], 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 (3), 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.

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

$$P\Big(X^{(r)}(t,d) + Z_{t,d,r} \in (x_{t,d,r} - \Delta x_{t,d,r}, x_{t,d,r} + \Delta x_{t,d,r}), t \in \mathcal{T}|\theta_{BD}(S)\Big) \tag{11}$$

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 (11) exactly requires summing over the possible sizes of the subpopulations at each time point, which is computationally infeasible. It is possible to speed up the computation using tools

from Hidden Markov Models (HMM), but even with these tools it is still computationally infeasible to compute the exact likelihood. We discuss this in further detail in S3 Text.

A more efficient approach is to use a Gaussian approximation. For the centered and normalized process $W_n(t)$ from (9), 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* $n \to \infty$,

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

where the (i, j) *element of the covariance matrix* Σ *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 S2 Text. Upon this proof, we further 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 (8):

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

The likelihood function is thus given by

$$L_{LC}(\theta_{BD}(S)|\mathbf{x}) = \prod_{r=1}^{N_R} \prod_{d \in \mathcal{D}} \frac{\exp\left[-\frac{1}{2}(\mathbf{x}_{\cdot,d,r} - \mu(d))^\top (n\Sigma(d) + c^2 I)^{-1} (\mathbf{x}_{\cdot,d,r} - \mu(d))\right]}{(\det(2\pi(n\Sigma(d) + c^2 I)))^{1/2}}. \quad (12)$$

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 (12) can improve the accuracy of parameter estimates, as we will discuss in the Results section. 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}.

The parameters used in our newly proposed models are summarized in Table 1:

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 (12), the distribution of the total cell number $\mathbf{X}(d) = \{X(t, d); t \in \mathcal{T}\}$ is approximated with a Gaussian distribution $N(\mu(d), n\Sigma(d) + c^2 I)$, whose mean and variance

Table 1. Table of parameters used in the end-points and live cell image methods.

Parameters	Description	Range
p_i	Initial proportion of each subpopulation	$\sum_{i=1}^{s} p_i = 1, p_i \geq 0$
β_i	Birth rate of subpopulation i	$\beta_i \in (0, 1)$
v_i	Death rate of subpopulation i	$v_i \in (max(0, \beta_{i-0.1}), \beta_i)$
b_i	Hill coefficient	$b_i \in (0, 1)$
E_i	Hill coefficient	$E_i > 0$
m_i	Hill coefficient	$m_i > 0$
n	Initial cell count	$\sim 10^4$
c	Standard deviation of observation noise	$(0, 0.15n)$

https://doi.org/10.1371/journal.pcbi.1011888.t001

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 $N(\mu(d), n\Sigma(d) + c^2 I)$ using a well-known measure of the distance between two distributions. The energy distance, introduced in [14], is a measure of the distance between probability distributions, which has previously been shown to be related to Cramer's distance [14, 15]. The energy distance has been utilized in several statistical tests [16] and is easily computed for multivariate distributions. For probability distributions F and G on $\mathbb{R}^d$, we define their energy distance as

$$D(F, G) = \sqrt{2\mathbb{E}[\|X - Y\|] - \mathbb{E}[\|X - X'\|] - \mathbb{E}[\|Y - Y'\|]}, \tag{13}$$

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 Eq (13) directly, we approximate the true energy distance by computing the empirical energy distance. For two sets of i.i.d. realization $\{X_1, \ldots, X_k\}$, $X_i \sim F$, $\{Y_1, \ldots, Y_m\}$, $Y_i \sim G$, one can obtain the empirical energy distance by

$$D_E(F, G) = \sqrt{\frac{2}{km}\sum_{i=1}^{k}\sum_{j=1}^{m}\|X_i - Y_j\| - \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\|X_i - X_j\| - \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\|Y_i - Y_j\|}. \tag{14}$$

Denote the distribution of $\mathbf{X}(d)$ by F_{BD} and the normal distribution $N(\mu(d), n\Sigma(d) + c^2 I)$ 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 (14). In Fig 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.

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 (10), and the more complex live cell imaging procedure ("live cell image method"), based on the likelihood L_{LC} in (12). 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. $x_{0,d,r} = n$. In the *in silico* experiment, n is always set to 1000, while in the *in vitro* experiment, it varies between experiments but is usually close to 3000.

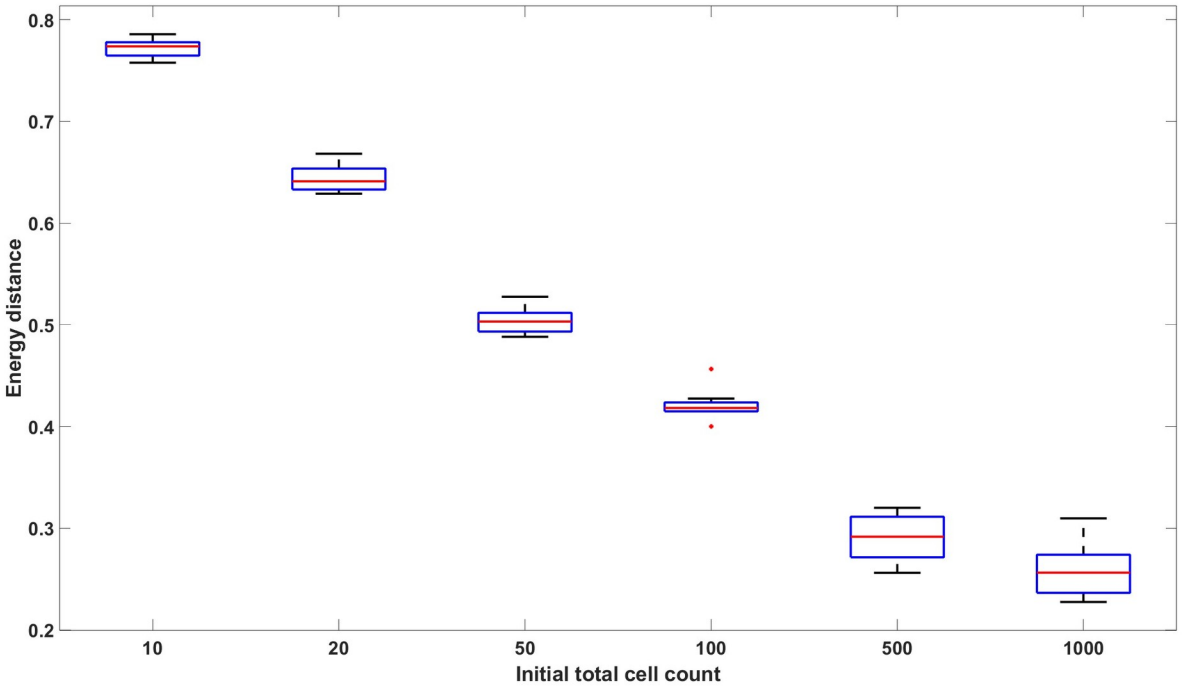

Fig 2. Empirical energy distance between linear birth-death simulated data and multivariate normal distributed data with respect to varying initial cell count: [10, 20, 50, 100, 500, 1000]. The data consists of $N_R = 100,000$ replicates and 7 time points $\mathcal{T} = [1, 2, 3, 4, 5, 6, 7]$. No drug 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.

https://doi.org/10.1371/journal.pcbi.1011888.g002

Application to simulated data

We first apply our estimation methods to simulated (*in silico*) data. In S1 Text, we provide details of the data generation and the parameter estimation for these *in silico* experiments. Data for this section are available at https://github.com/chenyuwu233/PhenoPop_stochastic/tree/main/In%20silico%20experiment.

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 2. 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. We furthermore assume that the data is collected at the time points

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

As in [8], 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

Table 2. Range for parameter generation of experiments with 2 subpopulations.

	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]	[0.5, 2.5]	[1.5, 5]	[0, 10]

https://doi.org/10.1371/journal.pcbi.1011888.t002

effect on the cell death rate, as is further explained in S1 Text. 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 S1 Text. The results of the case study are shown in Fig 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 $\theta^i_{BD}(2)$ for $i = 1, \ldots, 30$ is sampled from the ranges in Table 2. 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,

$$Er(\hat{x}; x) = \left|\log\left(\frac{x}{\hat{x}}\right)\right|. \tag{16}$$

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

In Fig 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

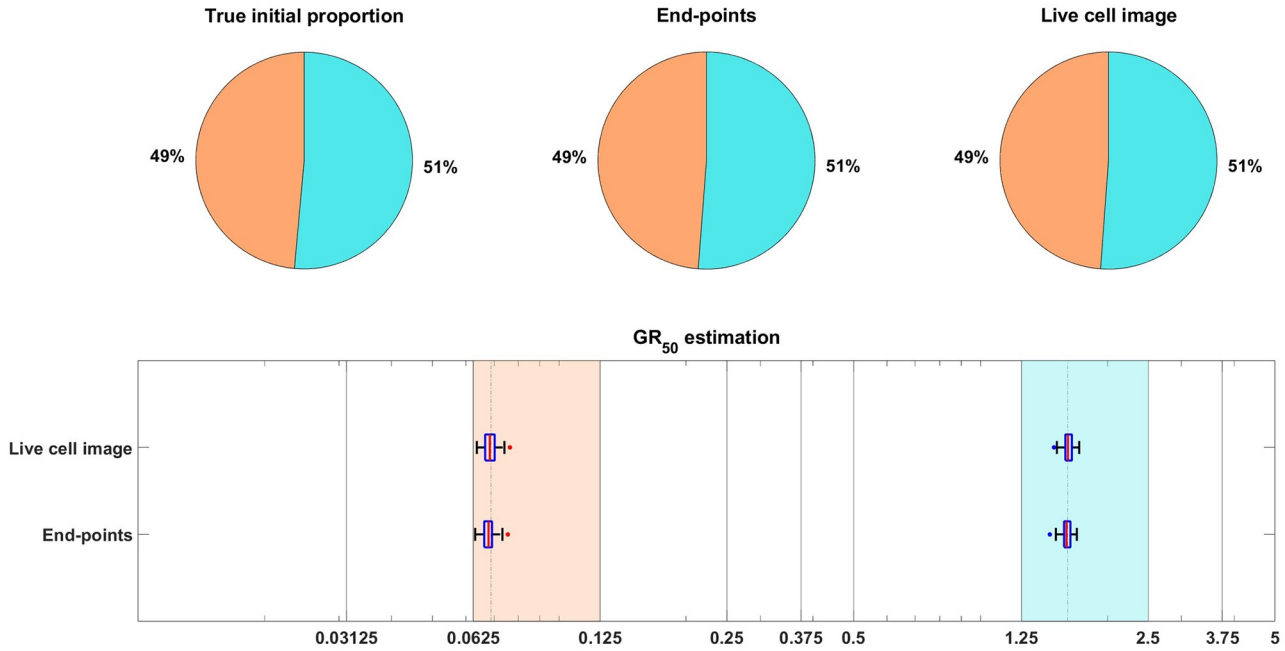

Fig 3. Estimation of the initial proportion and GR_{50} for 2 subpopulations using the end-points method and the live cell image method on simulated data. The parameter vector $\theta_{BD}(2)$ and observation noise c used in this example are $p_s = 0.4856$, $\beta_s = 0.1163$, $\nu_s = 0.0176$, $b_s = 0.8262$, $E_s = 0.0674$, $m_s = 4.5404$, $p_r = 0.5144$, $\beta_r = 0.4624$, $\nu_r = 0.3978$, $b_r = 0.8062$, $E_r = 1.5776$, $m_r = 4.2002$, $c = 1.2103$. 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, colored according to the corresponding colors, indicate the concentration intervals where the true sensitive GR_{50} and resistant GR_{50} are situated. 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.

https://doi.org/10.1371/journal.pcbi.1011888.g003

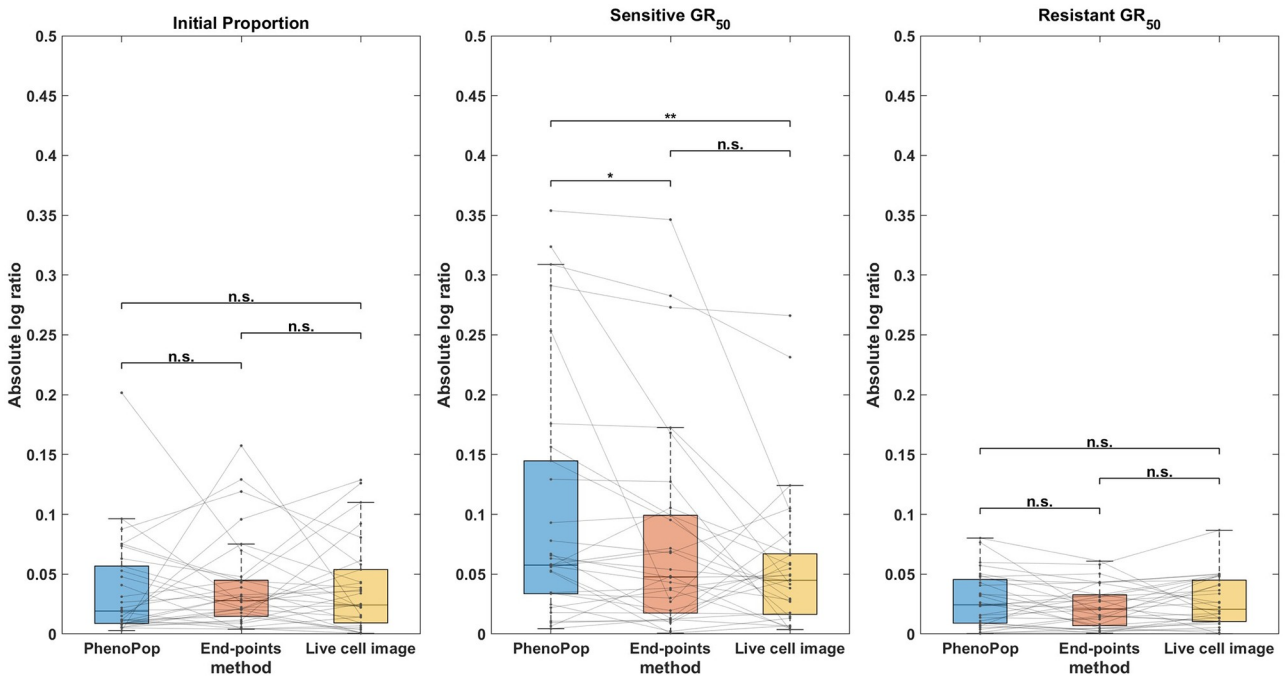

Fig 4. Absolute log ratio accuracy of three estimators $\{\hat{p}_s, \hat{GR}_s, \hat{GR}_r\}$ using the PhenoPop, end-points and live cell image methods. The results are summarized based on 30 different simulated datasets. This figure demonstrates that there are no significant differences in estimation accuracy among these three methods when the true parameters fall within the range described in Table 2.

https://doi.org/10.1371/journal.pcbi.1011888.g004

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. We have noticed that the CI width of the estimation depends on the true parameter value, especially for the GR_{50}. This becomes evident when comparing the CI widths for estimating the sensitive GR_{50} and resistant GR_{50}, as shown in Fig 5. Consequently, we have chosen the Wilcoxon signed rank test to illustrate the paired difference in the CI width between the three methods. By using paired differences, we can control for variability in CI width between different parameter sets.

In Fig 5, 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.

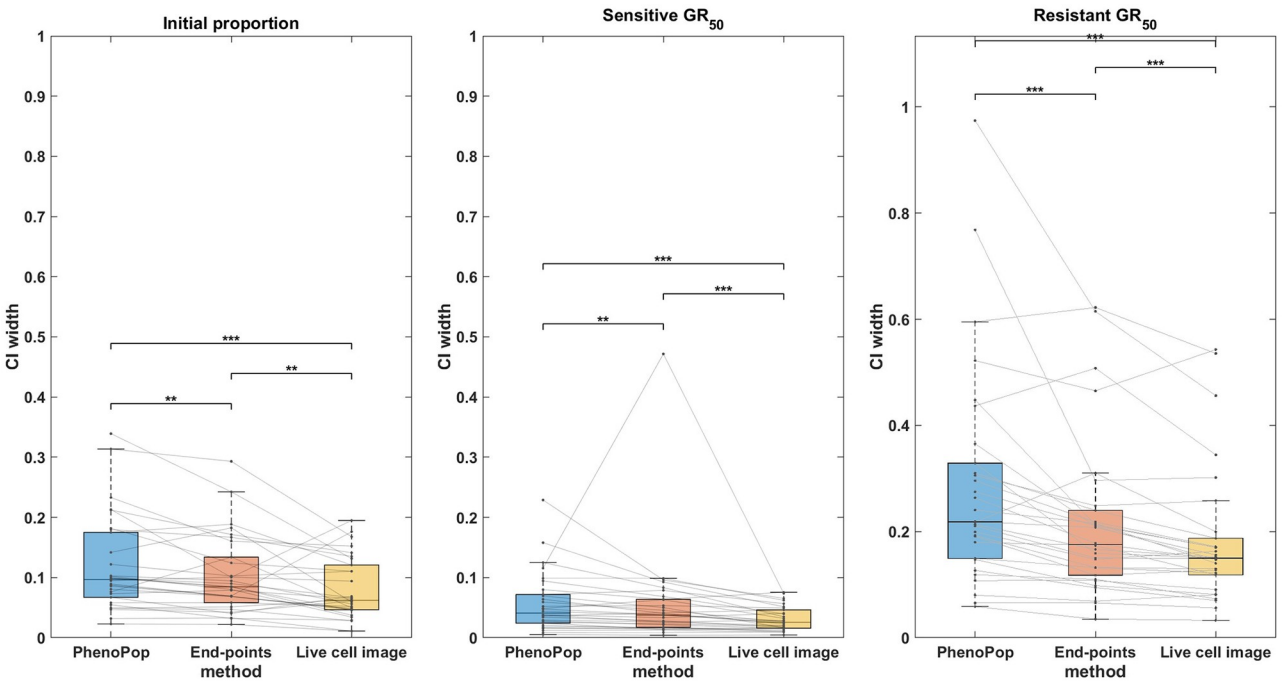

Fig 5. Comparison of the CI widths of three estimators $\{\hat{p}_s, \hat{GR}_s, \hat{GR}_r\}$ estimated from three different methods. The y-axis represents the CI width. The box plots summarize the results across 30 different simulated datasets. The significance bar indicates the p-values derived from the Wilcoxon signed rank test, with significance levels denoted as $^{***} \leq 0.001 \leq {}^{**} \leq 0.01 \leq {}^{*} \leq 0.05$. The solid line between the box plots indicates the paired result from all three methods. 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.

https://doi.org/10.1371/journal.pcbi.1011888.g005

Estimation of complete parameter set. Until now we have focused on estimating the initial fractions and GR_{50} values. Of course our model has several more parameters, and it is of interest how well the presented algorithms can estimate the complete parameter set, i.e. $\theta_{BD}(2) = \{(p_i, \beta_i, \nu_i, b_i, E_i, m_i), c; i = 1, 2\}$.

Based on the aforementioned 30 experiment results, we will investigate how accurately our algorithms can estimate the full parameter set θ_{BD}. In particular, we will look at the relative error

$$Er(\hat{x}; x) = \left| \frac{x - \hat{x}}{x} \right|$$

between the point estimate $\hat{x}$ and the true value x for each element of the parameter set.

In Fig 6, it is evident that the estimates of the initial fractions p and drug effect parameters (b, E) are accurate across all 30 *in silico* experiments. However, the estimates of m are not as accurate as those of the other parameters due to the fact that m introduces non-convexity to the optimization problem. Interestingly, our newly proposed models demonstrate the ability to reasonably recover most of the birth and death parameters (β, ν) of each subpopulation, which cannot be estimated using the PhenoPop method. This highlights a unique property of our newly proposed models, as they leverage data variability to infer the birth and death rates separately. Additionally, we observe as in Fig 4 that the parameters related to the sensitive subpopulation tend to be less accurately estimated than for the resistant subpopulation due to a smaller initial sensitive subpopulation proportion.

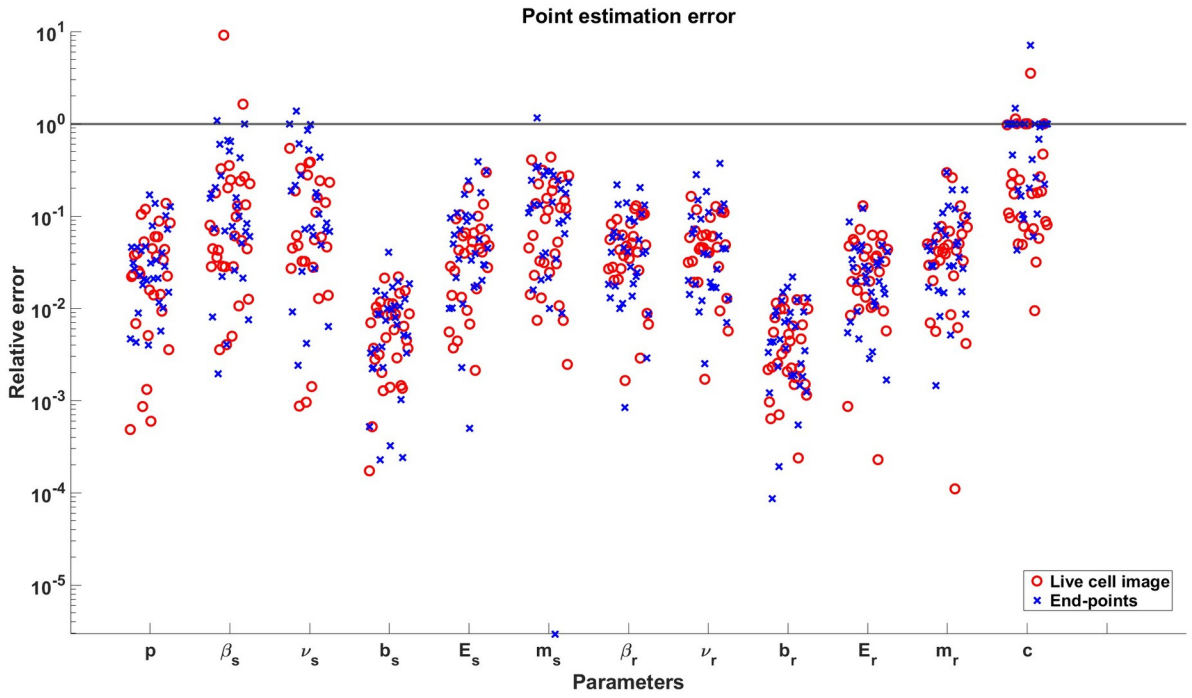

Fig 6. Relative point estimation errors for all parameters in the parameter vector θ_{BD}(2). Results from the live cell image method and the end-points method are included as shown in the legend. The y-axis is in logarithmic scale and the solid line indicates the place where the relative error is equal to 1.

https://doi.org/10.1371/journal.pcbi.1011888.g006

To analyze the trade-off in accuracy between estimates of different parameters, we proceeded to examine the empirical joint confidence region (JCR) for these parameters. Since we have 30 different sets of true parameters, we generate the empirical joint confidence region based on the estimation error, i.e. the difference between the estimate and the true parameter. Specifically, we calculated the errors for the 100 bootstrapped estimates for each experiment and compiled the results from the 30 experiments. Due to a discernible positive correlation between the estimates of the birth rate β and death rate ν, our focus in Fig 7 is solely on the errors associated with the parameters (β, b, E, m) for the two subpopulations. While we do not observe a clear trade-off in estimating (β, b, m) for both subpopulations, there is a slight hyperbolic shape in the JCR for the drug effect parameter E. This indicates that a smaller estimation error for E_s may cause a larger estimation error for E_r.

Effect of small time observation window on estimation accuracy. It should be noted that the quality of estimation depends on the quality of the experimental design, specifically the choice of time points and dosage levels. For instance, as we shorten the time window, the initial fluctuations in the linear birth-death process may make it difficult to infer and decouple the growth characteristics of the resistant and sensitive subpopulations. To illustrate how the three methods perform when only observing the first few hours of the experiment, we conducted a new set of experiments with time points:

$$\tau = \{0, 1/3, 2/3, 1, 4/3, 5/3, 2, 7/3, 8/3, 3, 10/3, 11/3, 4\}. \quad (17)$$

Otherwise, all the experimental settings are the same as the aforementioned experiment. In Fig 8, we compare the accuracy and precision in estimating the initial proportion between the PhenoPop and Live cell image methods. We find that both PhenoPop and Live cell image

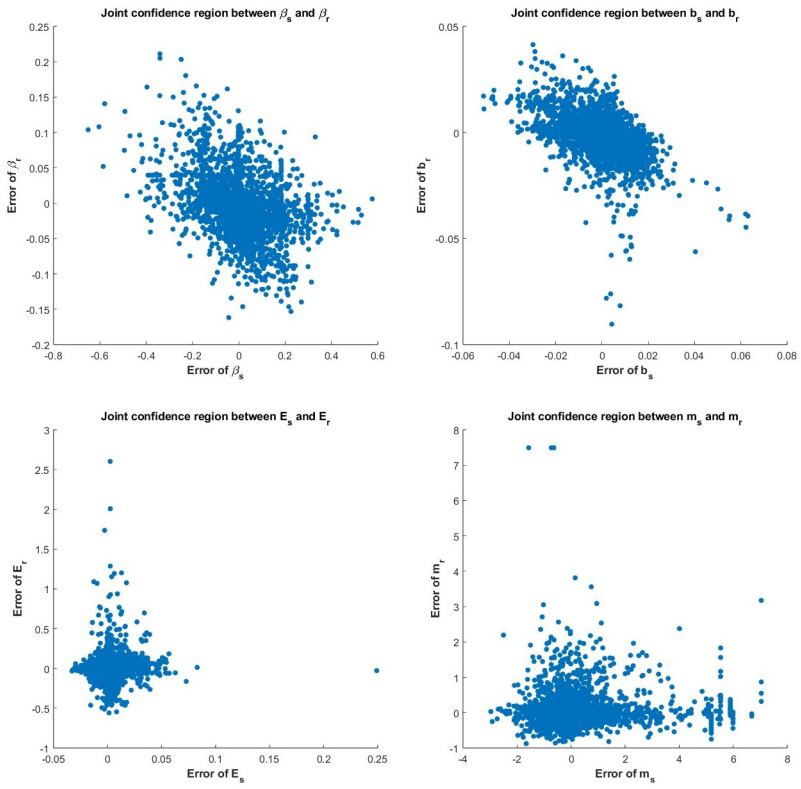

Fig 7. Joint confidence region of (β, b, E, m) between resistant and sensitive cells. In each subfigure, the x-axis represents the estimation error of the correspondent parameter of the sensitive subpopulation, while the y-axis represents the estimation error of the correspondent parameter of the resistant subpopulation.

https://doi.org/10.1371/journal.pcbi.1011888.g007

methods exhibit a significant deterioration in estimation accuracy and precision under this experimental design, which higlights the importance of collecting data over a sufficiently long time horizon.

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 3. Those parameters not listed in Table 3 are selected as in Table 2.

Fig 9 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.

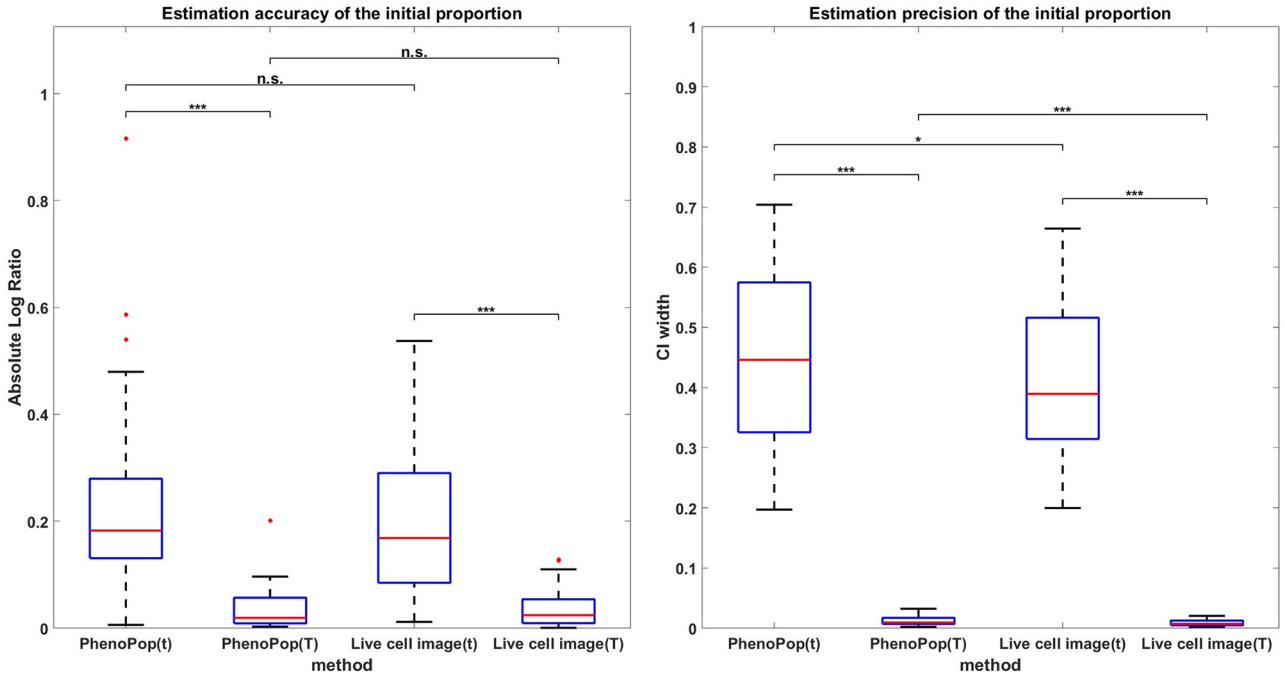

Fig 8. Estimation accuracy (left panel) and precision (right panel) of the initial proportion from data collected from different time horizons. The methods PhenoPop(t) and Live cell image(t) estimate from data collected at the time points τ defined in (17), while the methods PhenoPop(T) and Live cell image(T) estimate from data collected at the time points $\mathcal{T}$ defined in (15).

https://doi.org/10.1371/journal.pcbi.1011888.g008

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.

Performance in challenging conditions. In the previous work [8], 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.

Table 3. Modified range of parameters in experiments with 3 subpopulations.

	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]

https://doi.org/10.1371/journal.pcbi.1011888.t003

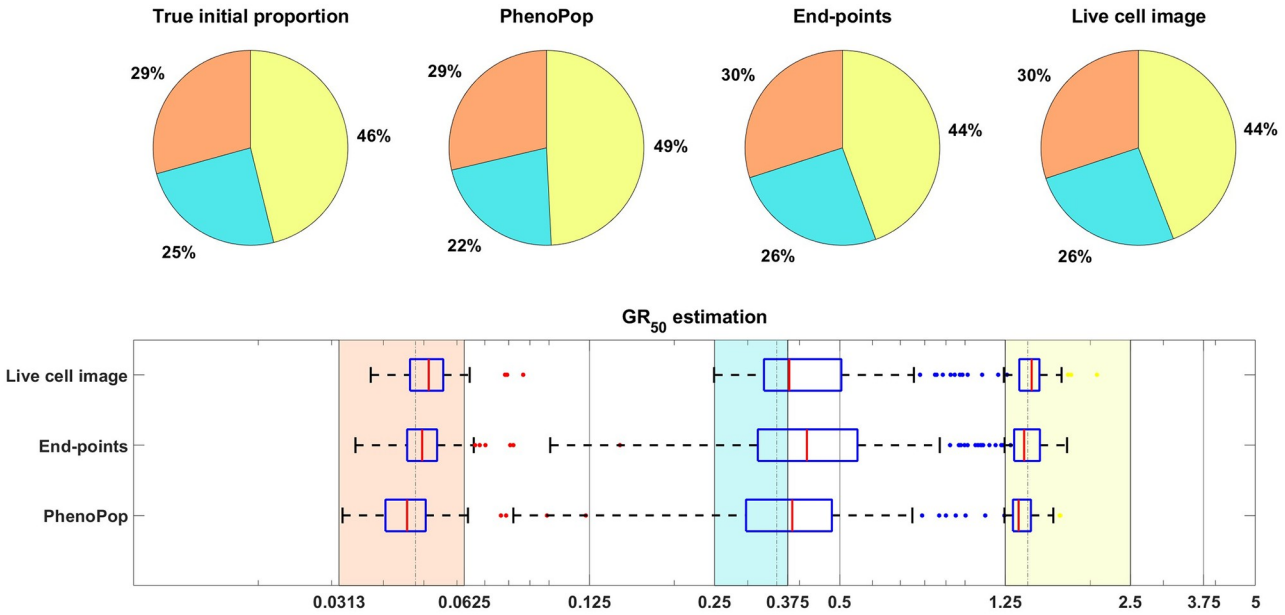

Fig 9. Estimation of the initial proportion and GR_{50} for 3 subpopulations using the three estimation methods. The parameter vector $\theta_{BD}(3)$ and the observation noise in this example are $p_s = 0.2135$, $\beta_s = 0.3214$, $\nu_s = 0.2773$, $b_s = 0.8782$, $E_s = 0.0344$, $m_s = 2.5998$, $p_m = 0.2718$, $\beta_m = 0.7334$, $\nu_m = 0.6776$, $b_m = 0.8506$, $E_m = 0.3558$, $m_m = 4.6055$, $p_r = 0.5147$, $\beta_r = 0.0683$, $\nu_r = 0.0253$, $b_r = 0.8614$, $E_r = 1.5764$, $m_r = 4.4706$, $c = 9.5209$. 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, colored according to the corresponding colors, indicate the concentration intervals where the true sensitive GR_{50} and resistant GR_{50} are situated. 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.

https://doi.org/10.1371/journal.pcbi.1011888.g009

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 2. The results are shown in Fig 10. 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 Fig 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 Fig 5 disappears under a very large observation noise.

We next evaluate estimation performance across 30 simulated datasets for each noise value $c \in \mathcal{C} = \{100, 200, 300, 400, 500\}$. Fig 11 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 Figs 12 and 13. For $c = 100$ (Fig 12), the precision advantage of the live cell image method over the other two methods is less pronounced than in Fig 5, where $c \in [0, 10]$, especially for the resistant GR_{50}. For $c = 500$ (Fig 13), the advantage in precision disappears for estimation of all three parameters. Note that both end-points and live-cell methods utilize information from the variability in the data to infer model parameters. Consequently, as observation noise increases, data variability becomes less informative, leading to the disappearance of the

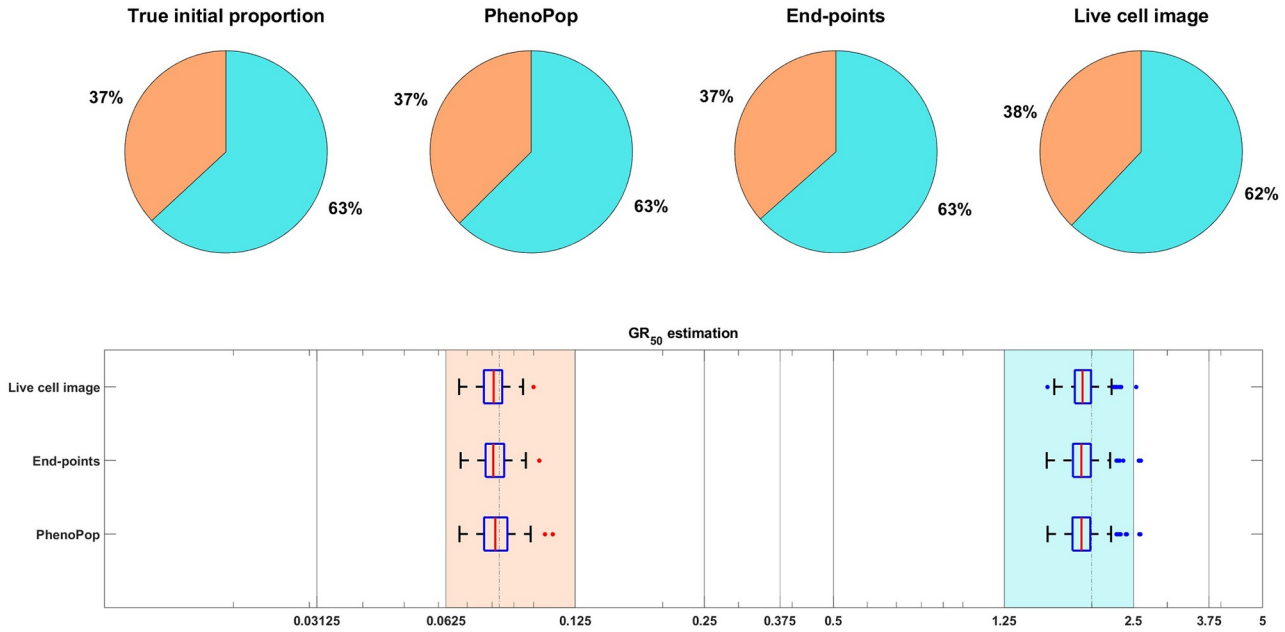

Fig 10. An illustrative example under the high observation noise scenario, i.e.$c = 500$. The parameter vector $\theta_{BD}(2)$ and the observation noise in this example are $p_s = 0.3690$, $\beta_s = 0.4380$, $\nu_s = 0.3422$, $b_s = 0.8398$, $E_s = 0.0813$, $m_s = 3.9647$, $p_r = 0.6310$, $\beta_r = 0.5320$, $\nu_r = 0.4767$, $b_r = 0.8674$, $E_r = 1.9793$, $m_r = 4.8357$, $c = 500$. Results are presented in Fig 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.

https://doi.org/10.1371/journal.pcbi.1011888.g010

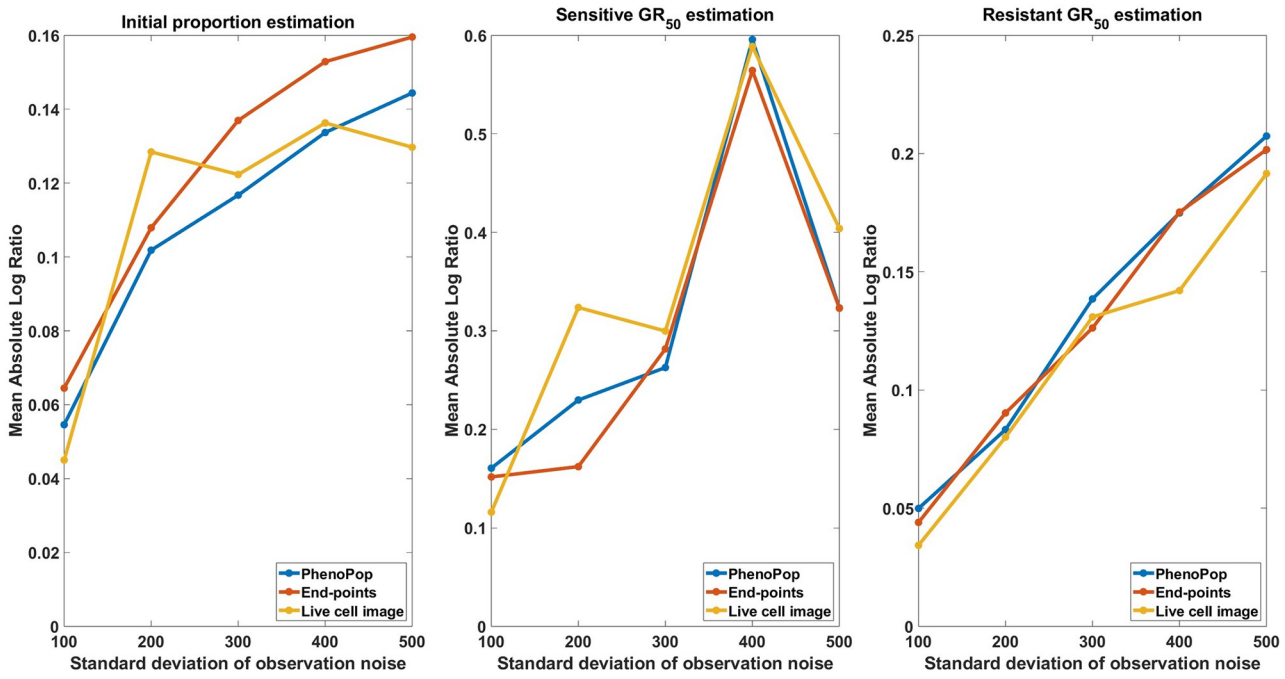

Fig 11. Estimation error of $\{\hat{p}_s, \hat{GR}_s, \hat{GR}_r\}$ with respect to varying standard deviation of observation noise. The metric of estimation error is the mean absolute log ratio of estimates across 30 simulated datasets, each generated from a distinct parameter set. The value of the 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 plots. Three different line plots correspond to three different methods, as indicated by the figure legends. This figure demonstrates that the estimates of the three methods deteriorate as the level of observation noise increases.

https://doi.org/10.1371/journal.pcbi.1011888.g011

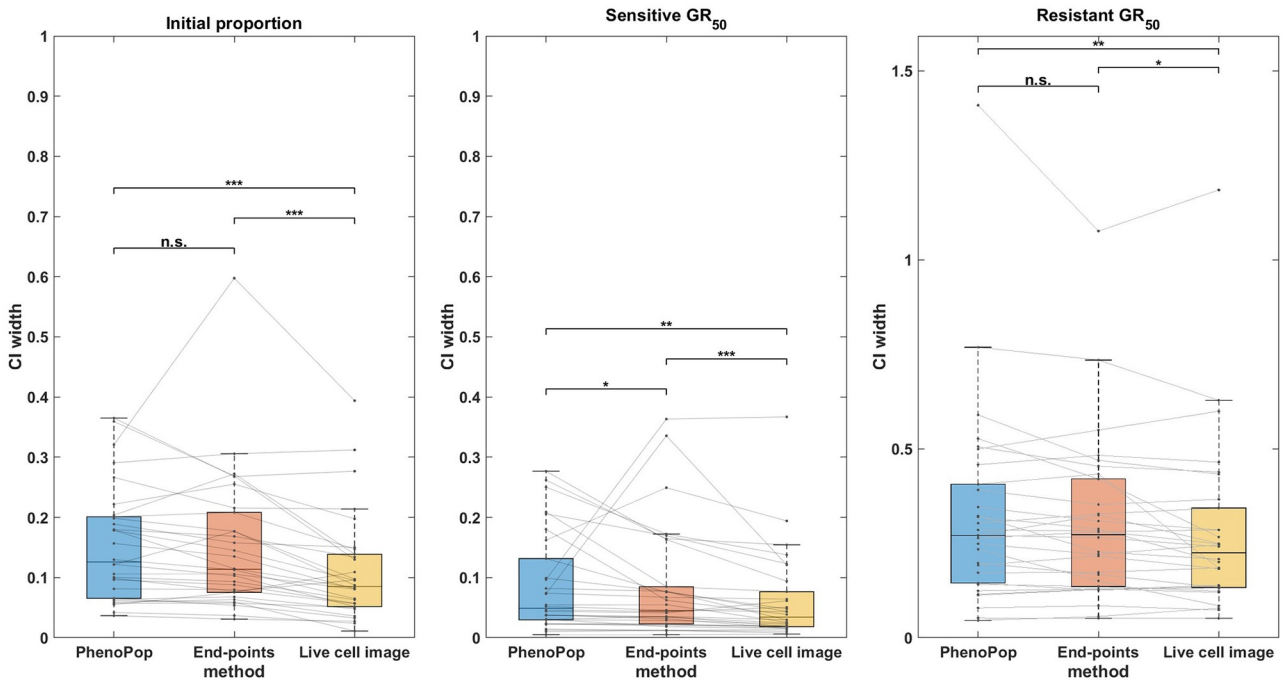

Fig 12. Comparison of the CI widths of the three estimators $\{\hat{p}_s, \hat{GR}_s, \hat{GR}_r\}$ using the three different estimation methods, when the observation noise parameter is set to $c = 100$. The y-axis represents the CI width. The box plot summarizes the results across 30 different datasets. 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 that 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.

https://doi.org/10.1371/journal.pcbi.1011888.g012

precision advantage offered by the newly proposed models. Importantly, however, Fig 12 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% [17, 18].

Small resistant subpopulation:

For the datasets investigated in Fig 3, 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 Fig 14, where p_s is assigned to 0.99, and other parameters are sampled according to Table 2. 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 Fig 15, 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

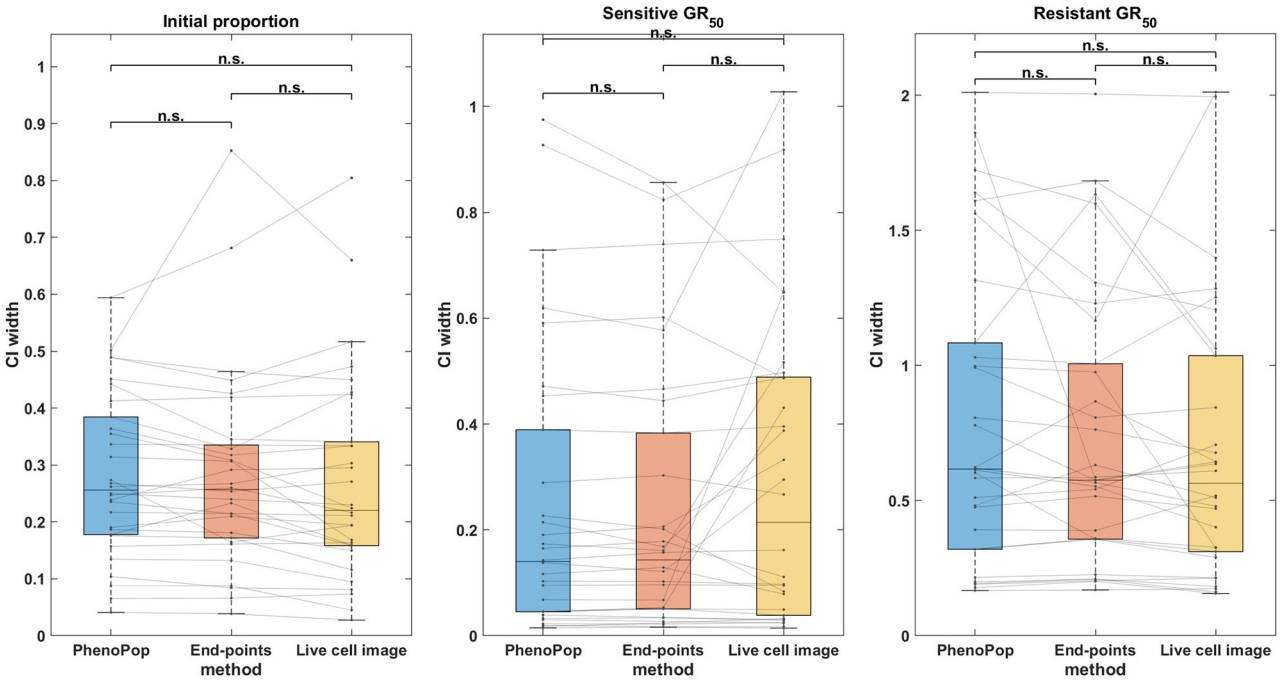

Fig 13. Comparison of the CI widths of three estimators $\{\hat{p}_s, \hat{GR}_s, \hat{GR}_r\}$ using the three different estimation methods, when the observation noise parameter is set to $c = 500$. Results are presented as in Fig 12. This figure demonstrates that the advantages of the live cell image method in estimation precision become less significant as the standard deviation of observation noise increases to 50% of the initial cell count.

https://doi.org/10.1371/journal.pcbi.1011888.g013

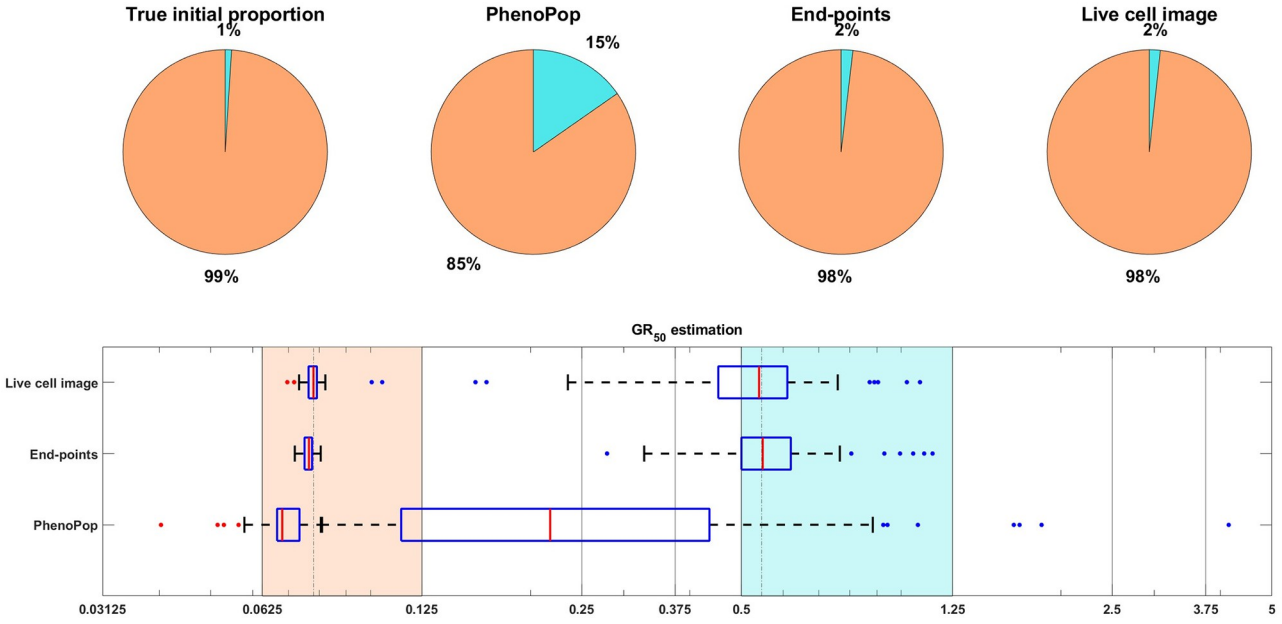

Fig 14. An illustrative example under the unbalanced initial proportion scenario, i.e. $p_s = 0.99$. The parameter vector $\theta_{BD}(2)$ and the observation noise in this example are $p_s = 0.9900$, $\beta_s = 0.4301$, $v_s = 0.4199$, $b_s = 0.8644$, $E_s = 0.0768$, $m_s = 4.3186$, $p_r = 0.0100$, $\beta_r = 0.1458$, $v_r = 0.1258$, $b_r = 0.8565$, $E_r = 0.5348$, $m_r = 3.7518$, $c = 4.8400$. Results are presented as in Fig 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.

https://doi.org/10.1371/journal.pcbi.1011888.g014

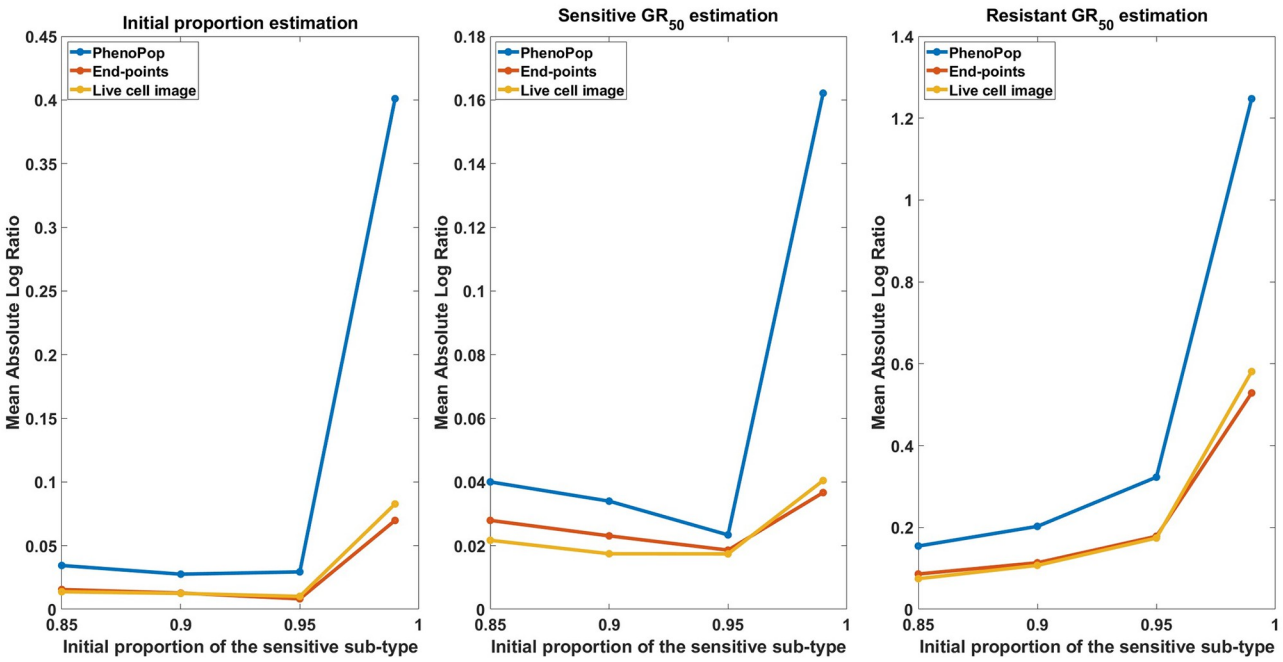

Fig 15. Estimation error of $\{\hat{p}_s, \hat{GR}_s, \hat{GR}_r\}$ with respect to varying resistant initial proportions. The metric of estimation error is the mean absolute log ratio across 100 simulated datasets, each generated from a distinct parameter set. The value of p_s in these 100 generating 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 plots. 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.

https://doi.org/10.1371/journal.pcbi.1011888.g015

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 Fig 4, where $p_s \in [0.3, 0.5]$. This further reinforces the hypothesis 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 Fig 3, 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. Fig 16 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 2. 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 Fig 17, 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 2, 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 GR_{50}, with all three methods showing similar degradation as the two subpopulations become more phenotypically similar.

Application to *in vitro* data

We conclude by evaluating the performance of our two new methods on *in vitro* experimental data. We provide the details of maximum likelihood estimation in S1 Text. Data for this

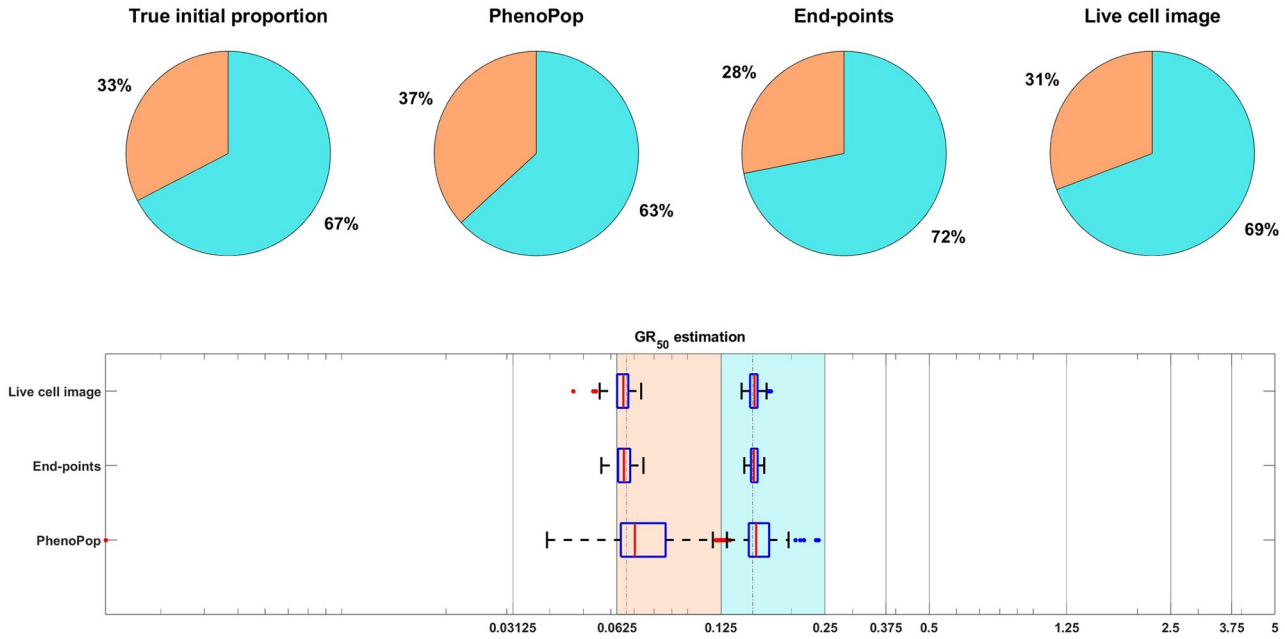

Fig 16. An illustrative example under the similar subpopulation sensitivity scenario. The parameter vector $\theta_{BD}(2)$ and the observation noise in this example are $p_s = 0.3263$, $\beta_s = 0.8896$, $\nu_s = 0.8215$, $b_s = 0.8820$, $E_s = 0.0654$, $m_s = 3.8539$, $p_r = 0.6737$, $\beta_r = 0.0925$, $\nu_r = 0.0661$, $b_r = 0.8171$, $E_r = 0.1500$, $m_r = 3.6015$, $c = 7.6660$. Results are presented as in Fig 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.

https://doi.org/10.1371/journal.pcbi.1011888.g016

section are available at https://github.com/chenyuwu233/PhenoPop_stochastic/tree/main/In%20vitro%20experiment.

The *in vitro* data consists of both monoclonal and mixtures of imatinib sensitive and resistant *Ba/F3* cells with different mixture proportions. 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. We gathered a total of four monoclonal datasets: two sensitive datasets and two resistant datasets, each initialized with different cell counts. These datasets are identified as SENSITIVE500, SENSITIVE1000, RESISTANT250, and RESISTANT500, with the numerical labels indicating the respective initial cell counts. We will also analyze mixture datasets with starting ratios between sensitive and resistant cells: 1 : 1, 1 : 2, 2 : 1 and 4 : 1. These datasets are denoted by BF11, BF12, BF21 and BF41, respectively. See [8] for further details on the experimental methods for generating the data.

In [8], we showed that the PhenoPop method can accurately identify the initial proportion of sensitive cells and both subpopulations' GR_{50} from the BF 21 and BF 41 datasets. Here, we employ our newly proposed methods to these two datasets and compare the result with the PhenoPop method. We present the point estimation results for the BF 41 and BF 21 datasets in Figs 18 and 19 respectively. It is important to note that we lack the true GR_{50} for the sensitive and resistant *Ba/F3* cells. Instead, we estimate the 'ground truth' GR_{50} based on independently applying all three methods to sensitive or resistant monoclonal data. The monoclonal estimates using the three methods indicate the likely range where the ground truth GR_{50} is located.

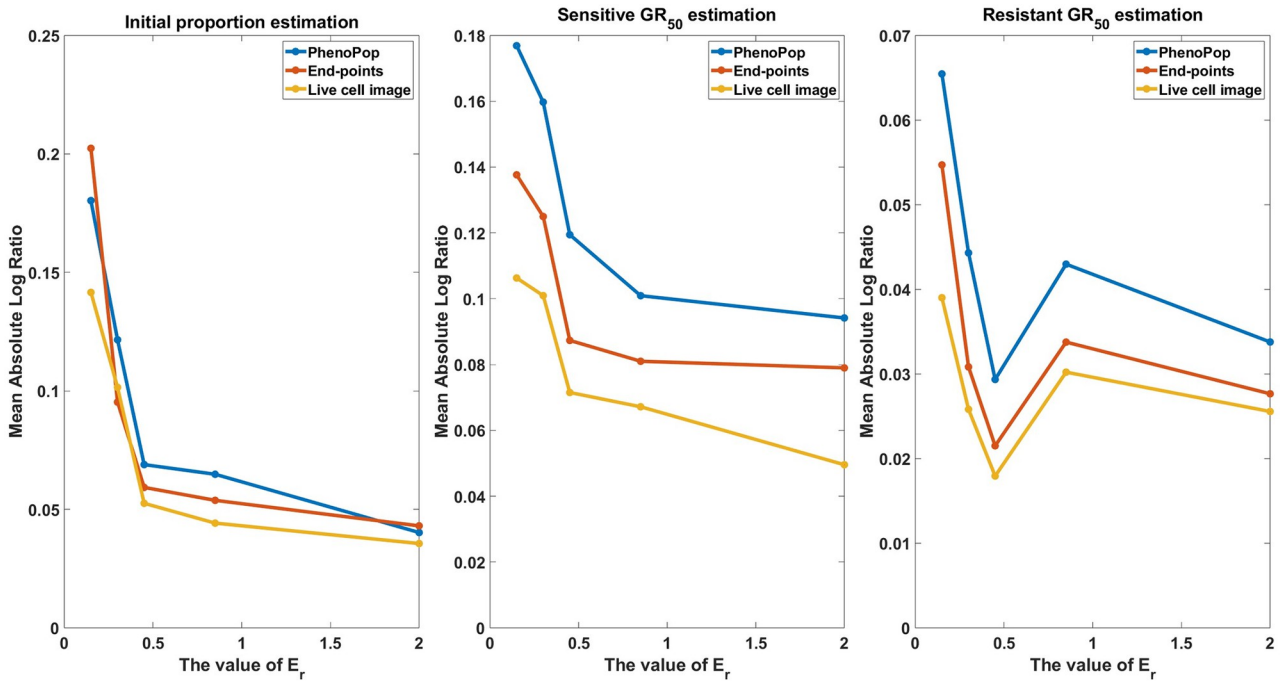

Fig 17. Estimation error of $\{\hat{p}_s, \hat{GR}_s, \hat{GR}_r\}$ with respect to varying similarity between subpopulation drug sensitivities. The metric of estimation error is the mean absolute log ratio across 80 simulated datasets, each generated from a distinct parameter set. The value of E_r in 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 plots. 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.

https://doi.org/10.1371/journal.pcbi.1011888.g017

For both the BF 41 and BF 21 datasets, and for both the sensitive and resistant subpopulation in each case, all three methods return GR_{50} estimates that fall within the same concentration interval, meaning that the estimates fall between two particular drug concentrations applied in the experiments (Figs 18 and 19). The estimates clearly indicate that the GR_{50}'s for the sensitive and resistant subpopulations are distinct, and the estimates furthermore are consistent with the likely 'ground truth' ranges for the GR_{50}'s. Finally, regarding subpopulation structure, there is a noticeable consensus among all three methods and the true initial proportions.

Next, we assess how well these three methods fit all the datasets using the Akaike Information Criterion (AIC). For a statistical model with parameters θ and likelihood function $\mathcal{L}(\theta|\mathbf{x})$, AIC is given by

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

Here, θ^* 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.

Results are shown in Table 4. 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 Material and methods, 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. Then, it is worth noting that the live cell image method has superior AIC scores among all three methods on the

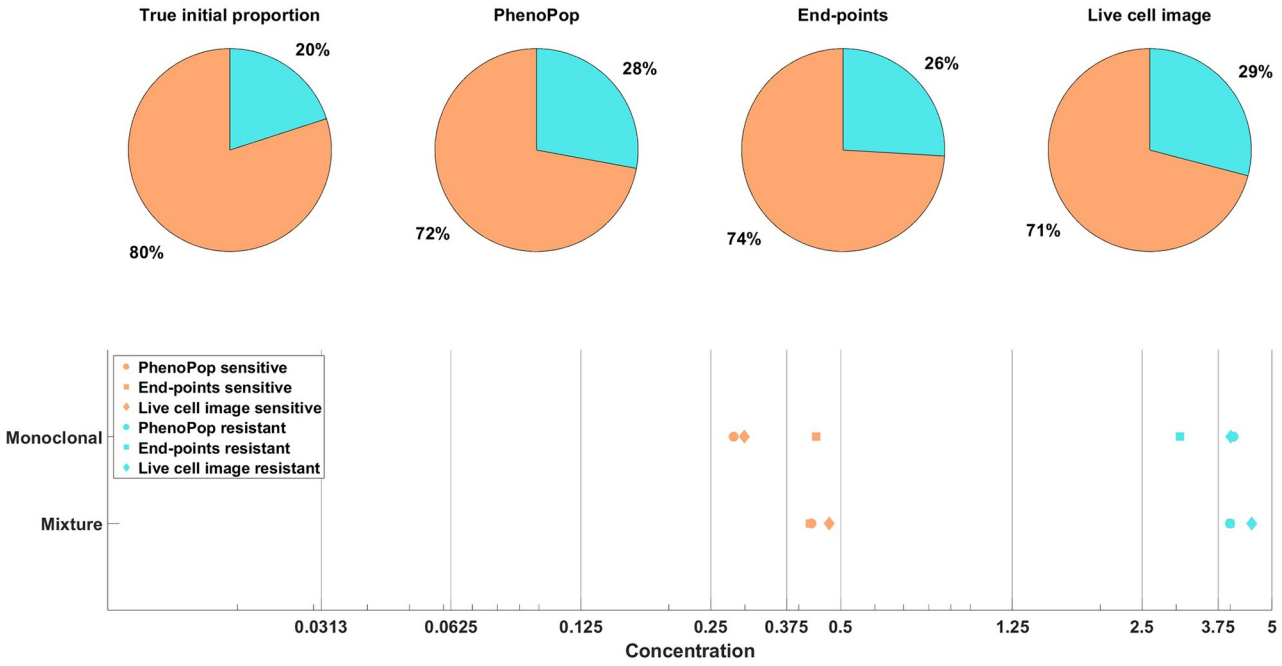

Fig 18. Estimation of the initial proportion and GR_{50} for 2 subpopulations from the mixture (BF41), and the monoclonal (SENSITIVE500, RESISTANT250) datasets using all three methods (PhenoPop, End-points, Live cell image). True initial proportion is obtained from the initial setting of the BF41 dataset, with a 4: 1 between sensitive and resistant cells. The estimated initial proportions are labeled according to their respective methods. True GR_{50} are estimated from monoclonal data and denoted as 'Monoclonal' from all three methods, while the estimated GR_{50} are labeled as 'Mixture'. Each shape in the GR_{50} estimation corresponds to a method, circle for PhenoPop, square for End-points, and diamond for Live cell image. Each color in the plot represents a distinct subpopulation: orange for sensitive and blue for resistant.

https://doi.org/10.1371/journal.pcbi.1011888.g018

monoclonal data. However, this advantage does not persist when fitting the mixture datasets. One possible explanation is that interactions within the mixture, which our model does not account for, may dilute the advantage of the live cell image method. Alternatively, the level of noise maybe high enough to reduce the power of the live cell method. Finally we observed that the auto-correlation structure of the data in the mixture experiments does not match the predicted structure of the live-cell model, while it does match the predicted structure in the monoclonal experiments. We plan to extend our models to account for potential cell-to-cell interactions in future work.

Conclusion

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 [8]. 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

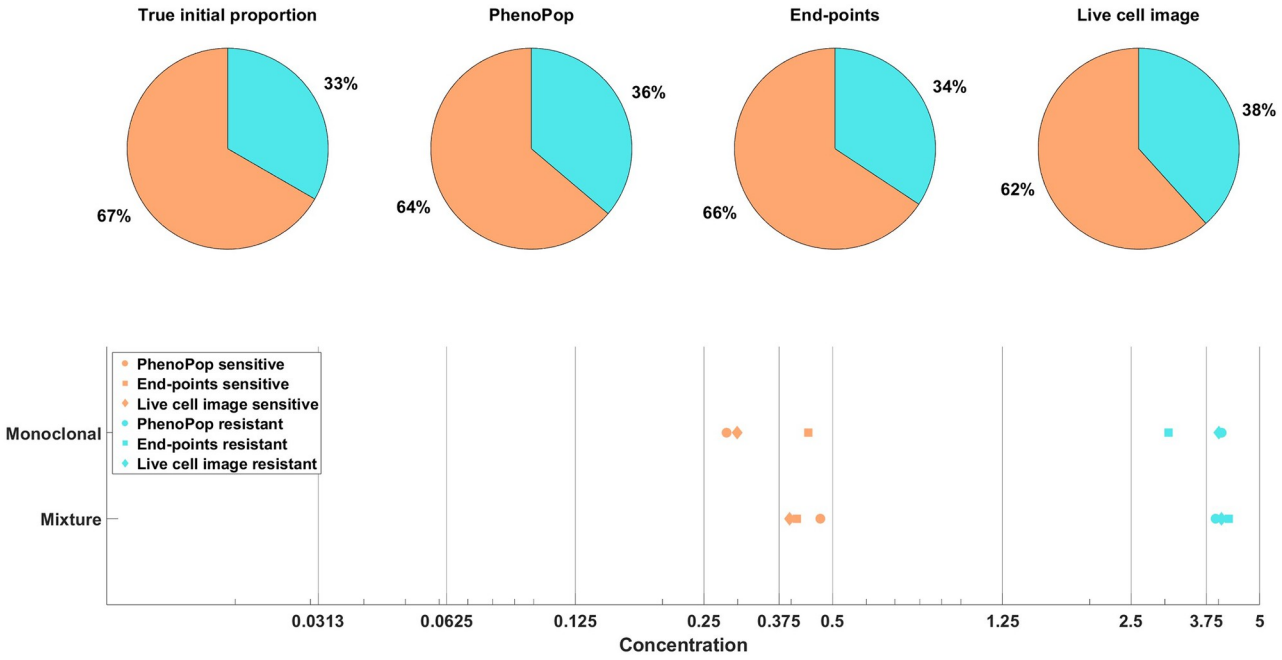

Fig 19. Estimation of the initial proportion and GR_{50} for 2 subpopulations from the mixture (BF21), and the monoclonal (SENSITIVE500, RESISTANT250) datasets using all three methods (PhenoPop, End-points, Live cell image). True initial proportion is obtained from the initial setting of the BF21 dataset, with a 2: 1 between sensitive and resistant cells. The estimated initial proportions are labeled according to their respective methods. True GR_{50} are estimated from monoclonal data and denoted as 'Monoclonal' from all three methods, while the estimated GR_{50} are labeled as 'Mixture'. Each shape in the GR_{50} estimation corresponds to a method, circle for PhenoPop, square for End-points, and diamond for Live cell image. Each color in the plot represents a distinct subpopulation: orange for sensitive and blue for resistant.

https://doi.org/10.1371/journal.pcbi.1011888.g019

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.

Table 4. AIC scores of three methods: PhenoPop method(PP), end-points method(EP), and live cell image method(LC) for the eight experimental datasets BF11, BF12, BF21, BF41, SENSITIVE500, SENSITIVE1000, RESISTANT250, RESISTANT500.

DATA	PP(AIC)	EP(AIC)	LC(AIC)
BF11	28502	26300	**25294**
BF12	30485	**26816**	27311
BF21	27928	**24064**	24182
BF41	28912	**24066**	24574
SENSITIVE500	16067	15776	**14021**
SENSITIVE1000	17977	16360	**15937**
RESISTANT250	15626	13053	**11353**
RESISTANT500	16886	12079	**11912**

https://doi.org/10.1371/journal.pcbi.1011888.t004

The second approach requires less cellular materials than the first approach when collecting data at multiple time points. In particular, when using live imaging techniques, we can obtain observations at multiple time points with a single sample, whereas when we use the end-point method multiple time points will require multiple independently run cell cultures. 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 [8], 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 supbopulations, 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 on estimating GR_{50} 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. These results show that the endpoints and live cell image methods are best used in datasets with lower noise standard deviation levels. Whereas when standard deviation levels are equal to a significant fraction of the initial population the precision benefit of these newer algorithms is reduced and one can safely use the original PhenoPop algorithm.

In our simulated datasets the statistical distribution of the data most closely matches the likelihood proposed by the live-cell imaging model. However, we see in many of our results that both the endpoints and PhenoPop algorithm are able to accurately identify the model parameters, with only a reduction in precision. We hypothesize that this is because both PhenoPop and the endpoints method accurately model the mean of the simulated data. As a result, they are able to use the mean behavior of the data to accurately identify the model parameters. Neither the endpoints nor PhenoPop method accurately models the variance of the data, and as a result they are not able to generate as precise estimates as the live-cell method on this data.

We finally compared the three methods using *in vitro* data. In particular, we used data from our previous work [8] 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. This may be due to cell-to-cell interactions, which we plan to incorporate into our models in future work. For monoclonal data, the live cell image method outperformed both PhenoPop and the end-points method in terms of model fit.

Similar to our previous work [8], our newly proposed methods effectively classify a heterogeneous tumor population into distinct phenotypic subpopulations with differentiable drug responses, using data on the response of the tumor bulk. Our methods identify the number of distinct subpopulations, thus bypassing the need for more advanced techniques like genotyping based on next-generation sequencing, and they simultaneously characterize the drug responses of each subpopulation, which would otherwise require separate assays. It is worth noting that our approach can also be applied to other controllable environmental factors, such as nutrient/oxygen deprivation. Accurately identifying subpopulations with differential drug sensitivities, especially small subpopulations with partial or full drug resistance, is crucial for the design of successful drug treatments. Indeed, drug response profiles for the tumor bulk may suggest treatment regimens which are effective at killing most of the tumor cells, while leaving a small reservoir of resistant cells with the ability to drive tumor recurrence. Knowing the subpopulation structure of the tumor can aid in the design of combination treatments which target several subpopulations simultaneously [19, 20]. In addition, the ability to infer the drug response profiles of each subpopulation facilitates the design of mathematically optimal drug treatments, under which drug administration schedules and dosage levels are chosen to maximize the probability of success [21, 22].

Beyond PhenoPop, the current work utilizes a linear birth-death process that allows for a more realistic variance model, particularly for experimental data obtained through live-cell imaging, where population size measurements are correlated over time. As a result of this more realistic variance model, our new inference methods show improved precision and accuracy over PhenoPop. An obvious future direction for this line of research is to leverage the inferred drug sensitivity profiles to design optimal treatments, as discussed above. In addition, there are several possible extensions and improvements of our statistical models, which can incorporate more complex cell population dynamics or capture important features of cell biology 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., [23, 24]). 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 ϵ, what observation set would allow us to identify the presence of this subpopulation? This is related to the broader question of parameter identifiability for our model, a question we plan to address in future work. Finally our stochastic model assumes that the time between cell divisions is exponential, but this is of course a great simplifcation. 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.

Supporting information

S1 Fig. Scatter plot of Residuals versus Time.
(TIF)

S1 Text. Details and extra results of the numerical experiments.
(PDF)

S2 Text. Proof and extension of proposition 2.
(PDF)

S3 Text. Exact path likelihood (11) computation.
(PDF)

Author Contributions

Conceptualization: Alvaro Köhn-Luque, Jasmine Foo, Kevin Leder.

Formal analysis: Chenyu Wu.

Funding acquisition: Jorrit Martijn Enserink, Arnoldo Frigessi, Kevin Leder.

Investigation: Chenyu Wu, Dagim Shiferaw Tadele, Kevin Leder.

Methodology: Einar Bjarki Gunnarsson, Even Moa Myklebust, Alvaro Köhn-Luque, Arnoldo Frigessi, Kevin Leder.

Supervision: Einar Bjarki Gunnarsson, Jorrit Martijn Enserink, Jasmine Foo, Kevin Leder.

Visualization: Chenyu Wu, Even Moa Myklebust, Alvaro Köhn-Luque, Dagim Shiferaw Tadele.

Writing – original draft: Chenyu Wu, Einar Bjarki Gunnarsson, Kevin Leder.

Writing – review & editing: Chenyu Wu, Einar Bjarki Gunnarsson, Even Moa Myklebust, Alvaro Köhn-Luque, Dagim Shiferaw Tadele, Jorrit Martijn Enserink, Arnoldo Frigessi, Jasmine Foo, Kevin Leder.

References

1. Pauli C, Hopkins BD, Prandi D, Shaw R, Fedrizzi T, Sboner A, et al. Personalized in vitro and in vivo cancer models to guide precision medicine. *Cancer discovery*, 7(5):462–477, 2017. https://doi.org/10.1158/2159-8290.CD-16-1154 PMID: 28331002
2. Grandori C and Kemp CJ. Personalized cancer models for target discovery and precision medicine. *Trends in cancer*, 4(9):634–642, 2018. https://doi.org/10.1016/j.trecan.2018.07.005 PMID: 30149881
3. Pemovska T, Kontro M, Yadav B, Edgren H, Eldfors S, Szwajda A, 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. https://doi.org/10.1158/2159-8290.CD-13-0350 PMID: 24056683
4. Pozdeyev N, Yoo M, Mackie R, Schweppe RE, Tan AC, and Haugen BR. Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies. *Oncotarget*, 7(32):51619, 2016. https://doi.org/10.18632/oncotarget.10010 PMID: 27322211
5. Matulis SM, Gupta VA, Neri P, Bahlis NJ, Maciag P, Leverson JD, et al. Functional profiling of venetoclax sensitivity can predict clinical response in multiple myeloma. *Leukemia*, 33(5):1291–1296, 2019. https://doi.org/10.1038/s41375-018-0374-8 PMID: 30679802
6. de Campos CB, Meurice N, Petit JL, Polito AN, Zhu YX, Wang P, et al. "direct to drug" screening as a precision medicine tool in multiple myeloma. *Blood cancer journal*, 10(5):1–16, 2020.
7. Marusyk A, Almendro V, and Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nature reviews cancer*, 12(5):323–334, 2012. https://doi.org/10.1038/nrc3261 PMID: 22513401

8. Köhn-Luque A, Myklebust EM, Tadele DS, Giliberto M, Schmiester L, Noory J, et al. Phenotypic deconvolution in heterogeneous cancer cell populations using drug-screening data. *Cell Reports Methods*, page 100417, 2023. https://doi.org/10.1016/j.crmeth.2023.100417 PMID: 37056380
9. Pakes AG. Ch. 18. biological applications of branching processes. *Handbook of statistics*, 21:693–773, 2003.
10. Kimmel M and Axelrod DE. *Branching processes in biology*. Springer, New York, 2002.
11. Durrett R. *Branching process models of cancer*. Springer International Publishing, 2014.
12. Tavaré S. The linear birth–death process: an inferential retrospective. *Advances in Applied Probability*, 50(A):253–269, 2018. https://doi.org/10.1017/apr.2018.84
13. Hannah R, Beck M, Moravec R, and Riss T. Celltiter-glo™ luminescent cell viability assay: a sensitive and rapid method for determining cell viability. *Promega Cell Notes*, 2:11–13, 2001.
14. Székely GJ. E-statistics: The energy of statistical samples. *Bowling Green State University, Department of Mathematics and Statistics Technical Report*, 3(05):1–18, 2003.
15. Cramér H. On the composition of elementary errors. *Scandinavian Actuarial Journal*, 1928(1):141–180, 1928. https://doi.org/10.1080/03461238.1928.10416872
16. Baringhaus L and Franz C. On a new multivariate two-sample test. *Journal of Multivariate Analysis*, 88 (1):190–206, 2004. https://doi.org/10.1016/S0047-259X(03)00079-4
17. Cadena-Herrera D, Esparza-De Lara JE, Ramírez-Ibañez ND, López-Morales CA, Pérez NO, Flores-Ortiz LF, et al. Validation of three viable-cell counting methods: Manual, semi-automated, and automated. *Biotechnology Reports*, 7:9–16, 2015. https://doi.org/10.1016/j.btre.2015.04.004 PMID: 28626709
18. Mumenthaler SM, Foo J, Leder K, Choi NC, Agus DB, Pao W, et al. Evolutionary modeling of combination treatment strategies to overcome resistance to tyrosine kinase inhibitors in non-small cell lung cancer. *Molecular pharmaceutics*, 8(6):2069–2079, 2011. https://doi.org/10.1021/mp200270v PMID: 21995722
19. He Q, Zhu J, Dingli D, Foo J, and Leder KZ. Optimized treatment schedules for chronic myeloid leukemia. *PLoS computational biology*, 12(10):e1005129, 2016. https://doi.org/10.1371/journal.pcbi.1005129 PMID: 27764087
20. Zhang J, Cunningham JJ, Brown JS, and Gatenby RA. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nature communications*, 8(1):1816, 2017. https://doi.org/10.1038/s41467-017-01968-5 PMID: 29180633
21. Marusyk A, Janiszewska M, and Polyak K. Intratumor heterogeneity: the rosetta stone of therapy resistance. *Cancer cell*, 37(4):471–484, 2020. https://doi.org/10.1016/j.ccell.2020.03.007 PMID: 32289271
22. Leder K, Pitter K, LaPlant Q, Hambardzumyan D, Ross BD, Chan TA, et al. Mathematical modeling of pdgf-driven glioblastoma reveals optimized radiation dosing schedules. *Cell*, 156(3):603–616, 2014. https://doi.org/10.1016/j.cell.2013.12.029 PMID: 24485463
23. Gunnarsson EB, De S, Leder K, and Foo J. Understanding the role of phenotypic switching in cancer drug resistance. *Journal of theoretical biology*, 490:110162, 2020. https://doi.org/10.1016/j.jtbi.2020.110162 PMID: 31953135
24. Gunnarsson EB, Foo J, and Leder K. Statistical inference of the rates of cell proliferation and phenotypic switching in cancer. *Journal of Theoretical Biology*, 568:111497, 2023. https://doi.org/10.1016/j.jtbi.2023.111497 PMID: 37087049