

RESEARCH ARTICLE

A multilevel hierarchical framework for quantification of experimental heterogeneity in population snapshot data

David J. Warne^{1*}, Xiangrun Zhu¹, Thomas P. Steele¹, Stuart T. Johnston², Scott A. Sisson³, Matthew Faria⁴, Ryan J. Murphy⁵, Alexander P. Browning^{2,6}

1 School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia, **2** School of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia, **3** School of Mathematics and Statistics, University of New South Wales, Sydney, Australia, **4** Department of Biomedical Engineering, The University of Melbourne, Melbourne, Australia, **5** School of Mathematical Sciences, Adelaide University, Adelaide, Australia, **6** Mathematical Institute, University of Oxford, Oxford, United Kingdom

* david.warne@qut.edu.au



OPEN ACCESS

Citation: Warne DJ, Zhu X, Steele TP, Johnston ST, Sisson SA, Faria M, et al. (2026) A multilevel hierarchical framework for quantification of experimental heterogeneity in population snapshot data. PLoS Comput Biol 22(6): e1014379. <https://doi.org/10.1371/journal.pcbi.1014379>

Editor: William Majoros, Duke University School of Medicine, UNITED STATES OF AMERICA

Received: January 28, 2026

Accepted: May 28, 2026

Published: June 15, 2026

Copyright: © 2026 Warne et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Code and relevant data to reproduce the results in this work are available on Git Hub at <https://github.com/davidwarne/multilevelHierarchicalODEs>.

Abstract

Biological systems exhibit substantial heterogeneity: that is, variation in specific characteristics of individuals within a population. As a result, it is of critical importance to appropriately account for biological heterogeneity when calibrating mathematical models to infer cellular processes and predict behaviour. Recent approaches consider ordinary differential equations with random parameters to quantify heterogeneity in dynamical processes of cells. In this setting, statistical inference is performed to characterise the distribution of these random parameters within a cell population. One significant limitation of this approach is the tacit assumption that there are no substantial deviations in these distributions across experimental replicates. In this work, we propose a flexible Bayesian hierarchical differential equation modelling framework that quantifies and distinguishes both inter-experimental heterogeneity (heterogeneity between experimental replicates) and intra-experimental heterogeneity (biological heterogeneity within replicate populations). We consider two recent studies that employ mathematical models to interpret flow cytometry snapshot data and quantify heterogeneity in nano-particle cell interactions and cell internalisation processes. Using simulation data, we demonstrate that substantial inaccuracy in the inferred dynamics can arise when experimental heterogeneity is not accounted for. By contrast, our hierarchical approach is robust to variability in inter-experimental and intra-experimental heterogeneity and our method simplifies to previous methods when inter-experimental heterogeneity is negligible. Our approach is flexible and widely applicable to applications involving replicate populations and snapshot data. We provide open-source implementations of our methods on [GitHub](https://github.com).

Funding: This project was supported by the Australian Research Council (ARC). DJW is supported by an ARC Discovery Early Career Researcher Award (DE250100396). STJ and MF are supported by an ARC Discovery Project (DP230100380). STJ is also supported by an ARC Discovery Project (DP250100477). SAS is supported by an ARC Discovery Project (DP260101134). DJW, APB, and STJ acknowledge support from the ARC Centre of Excellence for the Mathematical Analysis of Cellular Systems (MACSYS; CE230100001). APB thanks the Mathematical Institute, University of Oxford for a Hooke Research Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Author summary

In the design of personalised and targeted drugs and therapeutics it is essential to quantify the variability in the response of cellular processes involved in drug uptake. In practice, the experimental instruments and protocols used to collect cell population data for the study of such cellular processes are complex. As a result, it is important for statistical analyses that use these data to correctly account for variability between experiments in addition to the variability of cellular behaviour within the same experimental replicate. We present a flexible model of cellular dynamics that accounts for variability in biological, technical, and experimental replicates. We show that this leads to more accurate and robust results over standard approaches that pool replicates together.

Introduction

Accounting for biological heterogeneity is crucial in the biomedical sciences [1]. Heterogeneity plays a key role in the development and maintenance of healthy living organisms [2,3] through complex interactions between various types of cells driven by intracellular biochemical processes [4]. In pathological situations, such as cancer or viral infections [5–8], heterogeneity can seriously impact the efficacy of pharmaceutical interventions and eventual patient outcomes [9–11]. As a result, appropriately quantifying biological heterogeneity within a population of cells is of tremendous importance for understanding natural processes and for the design of effective drugs, especially in the areas of precision medicine such as targeted therapeutics [12–15].

A standard experimental approach to quantify properties of cell populations is flow cytometry [16–18]. In such an experiment, populations of cells (Fig 1(a)), are tagged or stained with fluorescent markers. The cells then individually flow through the cytometer nozzle and pass through a laser from which scattered light is amplified and converted to a digital signal representing a measurement of fluorescent intensity related to each cell in the population (Fig 1(b)). This results in a snapshot in time of the fluorescent intensity distribution across multiple colour channels, providing a rich source of information about potential variability in the cell populations at a single point in time (Fig 1(c)). Typically, thousands to tens of thousands of cells are analysed per time point.

Data obtained from flow cytometry experiments exhibit substantial levels of heterogeneity. While some of this variation can be attributed to measurement noise, it is well established that some of this variation arises due to intrinsic biological variation, even between genetically identical cells [19–23]. However, there are also extrinsic factors that drive heterogeneity in flow cytometry experiments that can adversely affect the statistical analysis of the biological heterogeneity of interest if not properly accounted for [24–26]. For example, subtle variation in sample preparation, instrument configuration for data acquisition, and manual user operation can have

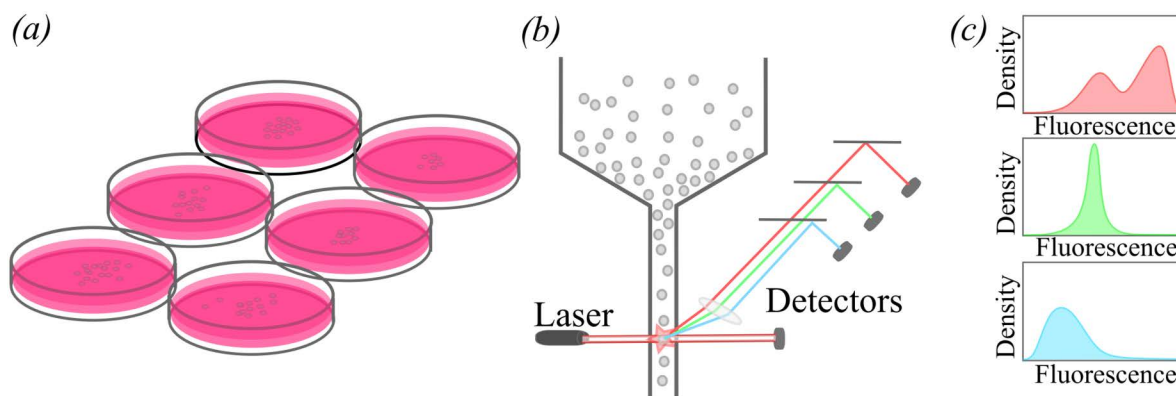


Fig 1. Schematic of flow cytometry data acquisition. (a) Cells are seeded into cell culture plates, e.g., 12-well plates, in a growth media (e.g., 1 mL) and incubated with various treatments for some time period (typically between 1 to 48 hours). (b) Cells are then removed from plates and placed in suspension, and passed through the flow cytometer nozzle to a laser beam where light scattering from the cell is measured at detectors. (c) For each fluorescent channel, the intensity distribution is obtained, leading to distributions showing cellular variability.

<https://doi.org/10.1371/journal.pcbi.1014379.g001>

a substantial effect on the signal-to-noise ratio and statistical bias in the data, potentially impacting the validity of conclusions drawn from the analysis [24,25,27,28].

In many studies, it is typically assumed that experimental replicates of identically prepared cell populations represent independent and identically distributed replicates. For example, in the work of Faria et al. [29], human leukaemia cells (THP-1 cell line) are seeded into a 12-well plate and nano-particle-cell interactions are observed via flow cytometry at six observation times (between 1 hr and 24 hrs), resulting in two replicates per observation time. In practice, statistical analysis is performed on the pooled sample of replicates [30,31], or each replicate is individually analysed [32,33]. Neither approach is ideal. Pooling ignores heterogeneity between experimental replicates and may lead to incorrect conclusions. Conversely, analysing replicates independently can lead to difficulties in assessing global trends [34]. A potential solution to analyse flow cytometry data is to use mathematical or statistical models that structure relationships between snapshots, and between time points data [34–38].

In this work, we introduce an approximate Bayesian multilevel hierarchical modelling framework for quantifying biological heterogeneity in flow cytometry data in situations where the parameters of interest are related to a mathematical model of cell dynamics based on differential equation models. The class of models we consider here is more general than previous work on non-linear mixed-effects models [35,36], as we allow for all cells to interact with a shared dynamic environment. To demonstrate our approach, we consider simulation studies based on two recent studies that quantify biological heterogeneity in cell behaviour using pooled flow cytometry data [39,40]. We highlight the dramatic impact that variation between replicates can have on pooled sample analysis results and show how a hierarchical approach can account for this effect. In addition, our approach reduces to the pooled case in situations where variation between replicates is minimal. Our approach is therefore generally applicable for applications in which controlling variation across flow cytometry replicates is infeasible.

Materials and methods

We develop a general Bayesian hierarchical modelling framework for the analysis of heterogeneity in cell behaviour using flow cytometry data. This heterogeneity is variability that can be intra-experimental heterogeneity within a replicate, or inter-experimental heterogeneity between replicates. This inter-experimental heterogeneity can be variation between

either technical replicates or experimental replicates. For clarity, we define a technical replicate to be an identically prepared replicate performed at the same time and the same person, and define an experimental replicate to be a replicate performed on a different day, and/or by a different person (See Fig 2). For example, if an experiment was performed with a 12-well plate where two-wells were assigned to each of six time points then this would represent two technical replicates. However, if the entire 12-well plate experiment was performed again at a different lab or with a different instrument then this would be two experimental replicates, each with two technical replicates. For most experimental protocols, we expect technical replicates to be closely related. As a result, we focus our attention to experimental replicates in this manuscript. However, we show how this can be extended to account for both technical and experimental replicates in S1 Appendix.

Flow cytometry data

We use simulated flow cytometry data that is a realistic characterisation of real flow cytometry data. Specifically, we build our data simulation process based on recent studies that investigate heterogeneity in interactions between cells and nano-particles [29,30,40] and cell internalisation processes [39,41,42].

Nano-engineered particle-cell interaction data. The flow cytometry data published by Faria et al., [29] and subsequently analysed for heterogeneity in particle-cell interaction by Murphy et al., [40] are for a human leukaemia monocytic suspension cell line, THP-1. Here, THP-1 cells are seeded in a 12-well plate and incubated with nano-particles (specifically, we focus on the 150 nm core-shell data) for 1, 2, 4, 8, 16, and 24 [hr]. After incubation, nano-particles that are not associated with cells are removed via washing. The nano-particle fluorescence across the cell population from each well plate is collected via flow cytometry. This results in snapshot data of associated nano-particle fluorescence for $N \approx 20,000$ cells for each incubation time point.

Dual-labelled probe internalisation data. Browning et al., [39] published flow cytometry data for a human B cell lymphoblast cell line (C1R) and analysed the heterogeneity in endocytosis of anti-transferrin (anti-TFR) antibodies. The measurement technique is based upon programmable sensors using DNA quenching probes [43,44]. Antibodies are dual-labelled with fluorescent dyes: FIP-Cy5 and BODIPY FL. Of these two probes, FIP-Cy5, is quenchable, that is, the

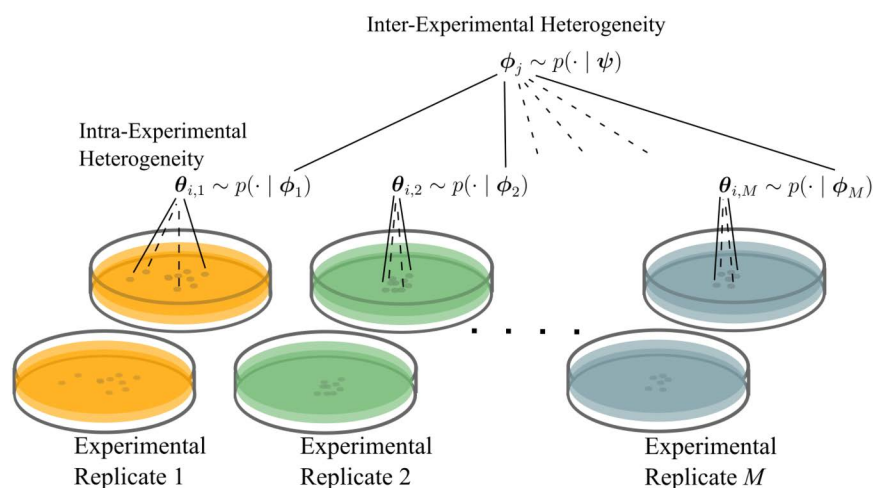


Fig 2. A schematic representation of the hierarchical statistical framework. Experimental replicates are indicated by different colours, with each experiment consisting of two identical technical replicates. Inter-experimental heterogeneity is captured through the hyper-parameters ψ that characterise the distribution of the M replicate specific hyper-parameters, $\phi_1, \phi_2, \dots, \phi_M$. Here, ϕ_j characterises the intra-experimental heterogeneity between N_j cells with θ_{ij} being the dynamic parameters for the i th cell within the j th replicate.

<https://doi.org/10.1371/journal.pcbi.1014379.g002>

fluorescence is effectively disabled on the cell surface using a quencher dye. Here, C1R cells are incubated with dual-labelled antibodies with incubation times of 5, 10, 20, 30, 60, 120, and 180 [min]. After incubation, cells are washed and resuspended both with and without the quencher [44]. The fluorescent signals for both FIP-Cy5 and BODIPY FL are measured for the cell populations in both quenched and unquenched samples via flow cytometry. This leads to snapshots of the jointly distributed FIP-Cy5 and BODIPY FL fluorescence for $N = 1,000$ cells over the incubation times and quenched/unquenched samples.

Mathematical modelling

We consider applications for which flow cytometry data is analysed using deterministic modelling based on ordinary differential equations (ODEs) as is commonplace in the mathematical and computational biology communities. The goal of such analysis is to quantify biological heterogeneity in parameters related to the dynamics of some cellular process. In this sense, each individual cell in the population is treated as having its own set of parameters and the goal is to find a suitable distribution for the parameters leading to cell dynamics matching the data distribution over time [39,40,45]. Here we describe the general ODE framework with random parameters used in the setting of pooled replicates, and then connect the framework to the particle-cell interaction model of Murphy et al., [40] and the cell internalisation model of Browning et al., [39]. These previous works focus exclusively on intra-experimental heterogeneity. We then show how to extend the analysis framework also account for inter-experimental heterogeneity between replicates and present a Bayesian multi-level hierarchical approach to quantify both intra- and inter-experimental heterogeneity together.

Cell population models with random parameters. We consider a representative and heterogeneous population of N cells. Let $\mathbf{X}_i$ be a vector of $d_X \geq 1$ dynamic properties of interest related to the i th cell (e.g., the average number of nanoparticles associated with cell, or concentrations of various biochemicals). This property vector evolves over time according to

$$\begin{aligned}\frac{d\mathbf{X}_i}{dt} &= f(\mathbf{X}_i(t), \mathbf{Z}(t), t; \theta_i), \\ \frac{d\mathbf{Z}}{dt} &= h(\mathcal{X}(t), \mathbf{Z}(t), t; \Theta_N),\end{aligned}\tag{1}$$

for $i = 1, 2, \dots, N$, where θ_i is a vector of $d_\theta \geq 1$ parameters with values specific to the i th cell, $\Theta_N = [\theta_1, \theta_2, \dots, \theta_N]$ is a $d_\theta \times N$ matrix of the population's parameters, $\mathcal{X}(t) = [\mathbf{X}_1(t), \mathbf{X}_2(t), \dots, \mathbf{X}_N(t)]$ is a $d_X \times N$ matrix representing the state of the cell population, and $\mathbf{Z}(t)$ is a vector of $d_Z \geq 1$ shared environmental variables that interact with cells in the population (e.g., a shared nutrient source). The functions $f(\mathbf{X}_i(t), \mathbf{Z}(t), t; \theta_i)$ and $h(\mathcal{X}(t), \mathbf{Z}(t), t; \Theta_N)$ define the dynamics of cells and the environment, respectively. In this setting, cells only indirectly interact via the shared environment. Direct interaction between cells could be incorporated in our framework by extending the cell dynamics to be a function of $\mathcal{X}(t)$ instead of $\mathbf{X}_i(t)$ only. For t_0 representing the start of the initial incubation time, we have initial conditions $\mathbf{X}_i(t_0) = \mathbf{x}_{i,0}$ for $i = 1, 2, \dots, N$, and $\mathbf{Z}(t_0) = \mathbf{z}_0$. Importantly, we do not assume that these initial conditions are known quantities, since such assumptions are known to cause problems in study reproducibility [46]. Instead, initial conditions are included as model parameters that we will estimate. This estimation need not be independent of the other parameters, but rather, initial conditions could be functionally dependent on the model parameters. For example, one may assume the environment or cell states to be in equilibrium that is determined based on a function of rate parameters.

The measured fluorescence of each cell is subject to external noise and thus treated as a random variable, that is,

$$\mathbf{Y}_i(t) \sim g(\cdot \mid \mathbf{X}_i(t), \theta_i, \eta),\tag{2}$$

where $\mathbf{Y}_i(t)$ are the recorded fluorescent intensities for the i th cell given the cell state $\mathbf{X}_i(t)$, dynamics parameters θ_i , and observation process specific parameters, η . The functional form of the observation process, $g(\cdot \mid \mathbf{X}_i(t), \theta_i, \eta)$, will typically

be determined based on calibration data for flow cytometry. The dimension of $\mathbf{Y}_i(t)$ will be the number of fluorescence channels and will generally not be the same as the dimension of $\mathbf{X}_i(t)$. Thus a single cell population snapshot obtained at time $t > t_0$ using flow cytometry is given by $\mathcal{Y}(t) = [\mathbf{Y}_1(t), \mathbf{Y}_2(t), \dots, \mathbf{Y}_N(t)]$. Typically, we do not observe the initial condition at $t = t_0$, and therefor it is treated as a latent variable to infer with the model parameters. Finally, for n observation times, $t_1 < t_2 < \dots < t_n$, we represent the entire flow cytometry dataset as $\mathcal{D} = [\mathcal{Y}(t_1), \mathcal{Y}(t_2), \dots, \mathcal{Y}(t_n)]$.

Since snapshot data as obtained through flow cytometry does not track individual cells, using such data to perform parameter inference on $\theta_1, \theta_2, \dots, \theta_N$ is not meaningful. Instead, one can consider $\theta_1, \theta_2, \dots, \theta_N$ to be independent identically distributed (i.i.d.) samples from a parametric distribution with probability density, $p(\theta | \phi)$, then perform parameter inference on the hyper-parameters $\phi \in \Phi$, where Φ is the hyper-parameter space. In a Bayesian setting, this leads to

$$p(\phi | \mathcal{D}) \propto p(\mathcal{D} | \phi)p(\phi),$$

where

$$p(\mathcal{D} | \phi) = \int_{\Theta^N} p(\mathcal{D} | \theta_1, \theta_2, \dots, \theta_N) \left[\prod_{i=1}^N p(\theta_i | \phi) d\theta_i \right].$$

This is the essence of the approach taken by Browning et al., [39] and Murphy et al., [40] and it relies on pooled snapshot data that requires the assumption of i.i.d. parameters across all cells and replicates. Our main contribution, that we present later, is an extension to this framework that only requires i.i.d. parameters within each experimental replicate.

Example 1: Particle-cell interaction model. The particle-cell interaction model of Murphy et al., [40], describes the association of free nano-particles to a population of cells. Given N cells the number of particles associated with the i th cell at time $t > t_0$, $P_i(t)$, is governed by the system,

$$\begin{aligned} \frac{dP_i}{dt} &= r_i c s \left(1 - \frac{P_i(t)}{K_i} \right) u(t), \\ \frac{du}{dt} &= -\frac{1}{N} \sum_{i=1}^N \frac{r_i c s}{v} \left(1 - \frac{P_i(t)}{K_i} \right) u(t), \end{aligned} \quad (3)$$

for $i = 1, 2, \dots, N$, where r_i and K_i are, respectively, the association rate and carrying capacity of particles for the i th cell, c is the fractional cell surface coverage, s is the surface area of the cell boundary, v is the volume of the well-mixed media, and $u(t)$ is the total free particle density with initial condition $u(t_0) = u_0$. In the context of our framework given by Eqs. (1) and (2), $\mathbf{X}_i(t) = P_i(t)$ and $\mathbf{Z}_i(t) = u(t)$ (i.e., $d_X = d_Z = 1$), and $\theta_i = [r_i, K_i]^T$ (i.e., $d_\theta = 2$).

The measured fluorescence, $Y_i(t)$, for cell i is modelled by

$$Y_i(t) = Y_{\text{cell},i} + (P_i(t) - \lfloor P_i(t) \rfloor) Y_{\text{particle}, \lceil P_i(t) \rceil} + \sum_{j=1}^{\lfloor P_i(t) \rfloor} Y_{\text{particle}, j},$$

where $Y_{\text{cell},i} \sim p(\cdot | \eta)$ is the autofluorescence of a cell, drawn at random from the empirical distribution cell-only calibration dataset, and $Y_{\text{particle},1}, \dots, Y_{\text{particle}, \lceil P_i(t) \rceil} \sim p_{\text{particle}}(\cdot | \eta)$ are individual particle fluorescences drawn at random from the empirical distribution particle-only calibration dataset. Here η are parameters associated with normalising voltages that are used to obtain flow cytometry measurements (See Murphy et al., [40] for details). Since there are $N=20,000$ cells, this leads to a computationally challenging system of ODEs as they are fully coupled through the environment (Eq. (3)). However, a tractable approximation with a semi-closed form solution can be obtained under the assumption that $u(t) \approx u_0$

throughout the simulation time (See Murphy et al., [40] for details). This assumption is quite common in the field and represents a high nano-particle dose such that the cells are not expected to meaningfully deplete it. Importantly, it does not lead to a completely decoupled system free from any environmental coupling and changes in $u(t)$ are still captured.

Example 2: Internalisation model. Browning et al., [39] consider a model describing the internalisation of transferrin antibodies that accounts for receptor recycling. Due to the experimental protocol (See Dual-labelled probe internalisation data), the concentration of free transferrin antibodies can be considered sufficiently high that unbound surface receptors are assumed to bind instantaneously to a free antibody. Furthermore, the high free antibody concentration enables the internalisation dynamics of all $N=2,000$ cells to be considered independent of each other since cells will never compete for free antibodies. For the i th cell at time $t > t_0$, Browning et al., [39] model the dynamics of the density of unbound internal receptors, $T_i(t)$, bound surface receptors, $S_i(t)$, internalised bound receptors, $E_i(t)$, and internalised free antibodies, $F_i(t)$, according to,

$$\begin{aligned}\frac{dT_i}{dt} &= -\beta_i T_i(t), \\ \frac{dS_i}{dt} &= \beta_i T_i(t) + p\beta E_i(t) - \lambda_i S_i(t), \\ \frac{dE_i}{dt} &= \lambda_i S_i(t) - p\beta E_i(t), \\ \frac{dF_i}{dt} &= -p\beta E_i(t),\end{aligned}\tag{4}$$

for $i = 1, 2, \dots, N$. Here, the initial total density of receptors for the i th cell is given by $T_i(t_0) + S_i(t_0) = R_i$ and it is assumed no antibodies have initially been internalised with $E_i(t_0) = F_i(t_0) = 0$. Further, $\lambda_i > 0$ and $\beta_i > 0$ are, respectively, the internalisation rate and recycling rate for the i th cell. Finally, $p \in (0, 1]$ is the disassociation probability that captures the probability that an internalised antibody is absorbed into the cell versus being recycled back to the surface. Following Browning et al., [39], p is associated with a purely chemical process and considered constant across cells, whereas λ_i , η_i and R_i are assumed to be heterogeneous across cells $i = 1, 2, \dots, N$. Relating this to our framework (Eqs. (1) and (2)), $\mathbf{X}_i(t) = [T_i(t), S_i(t), E_i(t), F_i(t)]^\top$ (i.e., $d_X = 4$), and $\theta_i = [\lambda_i, \beta_i, R_i, p]^\top$ (i.e., $d_\theta = 4$). In this model we do not have a dependent shared environment so $d_Z = 0$. In this case an analytic solution to Eq. (4) can be obtained using the matrix exponential (See Browning et al., [39] for details).

The observation process is more complex in this model than the particle-cell interaction model. Here, the experimental protocol involves dual-labelled antibodies with one of the labels (FIP-Cy5) being “quenchenable”, leading to a loss in FIP-Cy5 fluorescence for free or surface bound antibodies, and the other (BODIPY FL) being “unquenchenable” with fluorescence unaffected by the quenching process. The $N=2,000$ cells consist of two equal-sized technical replicates of size $\tilde{N} = 1,000$ with one replicate undergoing the *quenching* process before being processed through the flow cytometer. This leads to four measured fluorescence signal values for the i th pair of cells,

$$\mathbf{Y}_i(t) = [Q_i(t), U_i(t), \bar{Q}_i(t), \bar{U}_i(t)],$$

where $Q_i(t)$ (resp. $U_i(t)$) are the quenchenable FIP-Cy5 (resp. unquenchenable BODIPY FL) fluorescence signals measured from the unquenched sample group, and $\bar{Q}_i(t)$ (resp. $\bar{U}_i(t)$) are the quenchenable FIP-Cy5 (resp. unquenchenable BODIPY FL) fluorescence signals measured from the quenched sample group. The resulting observation process is

$$\begin{aligned}Q_i(t) &\sim \mathcal{N}(\alpha_1 [T_i(t) + S_i(t) + E_i(t) + F_i(t)] + A_Q, \sigma_1), \\ U_i(t) &\sim \mathcal{N}(\alpha_2 [T_i(t) + S_i(t) + E_i(t) + F_i(t)] + A_U, \sigma_2), \\ \bar{Q}_i(t) &\sim \mathcal{N}(\alpha_1 [(1-q)\bar{S}_i(t) + \bar{E}_i(t) + \bar{F}_i(t)] + A_Q, \sigma_1), \\ \bar{U}_i(t) &\sim \mathcal{N}(\alpha_2 [\bar{T}_i(t) + \bar{S}_i(t) + \bar{E}_i(t) + \bar{F}_i(t)] + A_U, \sigma_2),\end{aligned}\tag{5}$$

for $i = 1, 2, \dots, \tilde{N}$, where α_1 and σ_1 (resp. α_2 and σ_2) are intensity and noise the fluorescence measurement for FIP-Cy5 (resp. BODIPY FL). Further, A_Q (resp. A_U) corresponds to the average autofluorescence of FIP-Cy5 (resp. BODIPY FL), and q is the quenching efficiency. While A_Q , A_U and q are pre-estimated from data directly, the parameters $\eta = (\alpha_1, \alpha_2, \sigma_1, \sigma_2)$ must be inferred with the model parameters.

Statistical framework

In the previous sections, we describe the typical setting in which biological heterogeneity has been treated in the literature. That is, replicates are pooled and treated as a single snapshot population [29–31,39,40]. In this work, we propose an extension in which each replicate is treated as its own population with variation within replicates (i.e., intra-experimental heterogeneity) being distinguished from variation between replicates (i.e., inter-experimental heterogeneity). To simplify the exposition of our approach we present a two-level Bayesian hierarchical model, however, the approach can be extended to more levels as shown in the [S1 Appendix](#).

Inter-experimental and intra-experimental heterogeneity. Within the random ODE framework ([Eqs. \(1\) and \(2\)](#)), we consider M experimental replicates with the j th replicate analysing N_j cells in total across identical technical replicates ([Fig 2](#)). The cell population dynamic model ([Eq. \(1\)](#)) becomes,

$$\begin{aligned}\frac{d\mathbf{X}_{ij}}{dt} &= f(\mathbf{X}_{ij}(t), \mathbf{Z}_j(t), t; \theta_{ij}), \\ \frac{d\mathbf{Z}_j}{dt} &= h(\mathcal{X}_j(t), \mathbf{Z}_j(t), t; \Theta_{N_j,j}),\end{aligned}\quad (6)$$

for $i = 1, 2, \dots, N_j$, and $j = 1, 2, \dots, M$. Here, $\mathbf{X}_{ij}(t)$ is the state of the i th cell in the j th replicate having parameters θ_{ij} . For the j th replicate we have the subpopulation state, $\mathcal{X}_j(t) = [\mathbf{X}_{1,j}(t), \mathbf{X}_{2,j}(t), \dots, \mathbf{X}_{N_j,j}(t)]$, the shared environment, $\mathbf{Z}_j(t)$, and the subpopulation parameters $\Theta_{N_j,j} = [\theta_{1,j}, \theta_{2,j}, \dots, \theta_{N_j,j}]$.

Similarly, the observation process ([Eq. \(2\)](#)) becomes,

$$\mathbf{Y}_{ij}(t) \sim g(\cdot | \mathbf{X}_{ij}(t), \theta_{ij}, \eta), \quad (7)$$

where $\mathbf{Y}_{ij}(t)$ is the fluorescence of the i th cell in the j th replicate. The snapshot data for the j th replicate at time $t > t_0$ is $\mathcal{Y}_j(t) = [\mathbf{Y}_{1,j}(t), \mathbf{Y}_{2,j}(t), \dots, \mathbf{Y}_{N_j,j}(t)]$. We assume that the groups of M replicates are observed at the same n observations times $t_1 < t_2 < \dots < t_n$, and denote $D_j = [\mathcal{Y}_j(t_1), \mathcal{Y}_j(t_2), \dots, \mathcal{Y}_j(t_n)]$ for the series of snapshots nominally labelled as the j th replicate, as in reality there are $n \times M$ replicates as the process of taking the snapshot typically destroys the sample. This leads to the complete flow cytometry dataset, including the set of snapshots in replicate subgroups, to be represented by $\mathcal{D} = [D_1, D_2, \dots, D_M]$.

Here, we assume that $\theta_{1,j}, \theta_{2,j}, \dots, \theta_{N_j,j}$ are i.i.d. from a parametric distribution $p(\theta | \phi_j)$ with subgroup hyper-parameter, ϕ_j , representing the distribution parameters for the j th replicate, this subgroup hyper-parameter may also include the number of observed cells N_j if it varies between replicates. Across all replicates, these hyper-parameters $\phi_1, \phi_2, \dots, \phi_M$, are distributed (i.i.d.) according to another parametric distribution, $p(\phi | \psi)$ where ψ are the between group hyper-parameters ([Fig 2](#)).

Computational inference. To quantify both inter-experimental and intra-experimental heterogeneity, characterised by $p(\cdot | \psi)$ and $p(\cdot | \phi_1), p(\cdot | \phi_2), \dots, p(\cdot | \phi_M)$, respectively, we aim to infer the hyper-parameters, $\psi, \phi_1, \phi_2, \dots, \phi_M$. That is,

$$p(\psi, \phi_1, \dots, \phi_M | \mathcal{D}) \propto p(\mathcal{D} | \psi, \phi_1, \dots, \phi_M) p(\psi, \phi_1, \dots, \phi_M), \quad (8)$$

where the joint prior that enforces the hierarchical structure of the model is given by

$$p(\psi, \phi_1, \dots, \phi_M) = p(\psi) \prod_{j=1}^M p(\phi_j | \psi), \quad (9)$$

and the likelihood is

$$p(\mathcal{D} | \psi, \phi_1, \dots, \phi_M) = \prod_{j=1}^M \left\{ \int_{\Theta^N} p(D_j | \Theta_{N,j}) p_j(\Theta_{N,j} | \phi_j) \left[\prod_{i=1}^{N_j} d\theta_{ij} \right] \right\}, \quad (10)$$

with $p_j(\Theta_{N,j} | \phi_j) = \prod_{i=1}^{N_j} p(\theta_{ij} | \phi_j)$. In Eq. (10) we integrate out the individual cell parameters, θ_{ij} , within the likelihood function. This leads to a complex likelihood evaluation while reducing the dimension of the inference problem substantially to the space of hyper-parameters $\Phi^M \times \Psi$.

To maintain the lower dimensional inference problem from Eqs. (8)–(10) while avoiding the the direct likelihood calculation (10), we adopt an Approximate Bayesian computation (ABC) approach [47–49]. That is, we approximate Eq. (8) using the ABC posterior,

$$p_{\text{ABC}}(\psi, \phi_1, \dots, \phi_M | \mathcal{D}) \propto \mathbb{P}(\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon | \phi_1, \dots, \phi_M) p(\phi_1, \dots, \phi_M | \psi) p(\psi). \quad (11)$$

where $\mathcal{D}_s$ is simulated data and $\rho(\mathcal{D}, \mathcal{D}_s)$ is a distribution matching discrepancy metric based on the Anderson-Darling distance,

$$\rho(\mathcal{D}, \mathcal{D}_s) = \frac{1}{M} \sum_{j=1}^M \sum_{k=1}^n A(\mathcal{Y}_j(t_k), \mathcal{Y}_j^s(t_k)), \quad (12)$$

where $\mathcal{Y}_j(t_k) = [Y_{1,j}(t_k), Y_{2,j}(t_k), \dots, Y_{N,j}(t_k)]$ is the real data snapshot for replicate j at time t_k , and $\mathcal{Y}_j^s(t_k) = [Y_{1,j}^s(t_k), Y_{2,j}^s(t_k), \dots, Y_{N,j}^s(t_k)]$ is the simulated data snapshot for replicate j at time t_k . Here, the function $A(\mathcal{X}, \mathcal{Y})$ is the Anderson-Darling distance,

$$A(\mathcal{X}, \mathcal{Y}) = \left[-N - \sum_{i=1}^N \frac{2i-1}{N} \log(\hat{F}_{\mathcal{X}}(Y_i)) + \log(1 - \hat{F}_{\mathcal{X}}(Y_{N+1-i})) \right]^{1/2}, \quad (13)$$

where $\mathcal{X} = \{X_1, X_2, \dots, X_N\}$ and $\mathcal{Y} = \{Y_1, Y_2, \dots, Y_N\}$ are two sample sets and $\hat{F}_{\mathcal{X}}(x)$ is the empirical cumulative distribution for $\mathcal{X}$. In this setting, we have that $p_{\text{ABC}}(\phi_1, \phi_2, \dots, \phi_M, \psi | \mathcal{D}) \rightarrow p(\psi, \phi_1, \phi_2, \dots, \phi_M | \mathcal{D})$ as $\epsilon \rightarrow 0$. We use an adaptive ABC scheme based on sequential Monte Carlo methods of Drovandi and Pettitt [50] that refines ϵ through sequential importance resampling (S2 Appendix).

Results

To demonstrate the advantages of our approach, we perform simulations that represent *in silico* repeats of the experiments described in the Flow cytometry data Section based on published data [29,39,40]. The simulation setting enables experimental heterogeneity to be directly controlled and results to be compared to a known ground truth.

Nano-engineered particle-cell interaction experiments

Murphy et al., [40] consider an ODE model of particle-cell interactions (See Example 1: Particle-cell interaction model) with each cell, $i = 1, 2, \dots, N$, having its own particle association rate, r_i , and carrying capacity, K_i . In the pooled sample setting, they consider these parameters $\theta_i = (r_i, K_i)$ to be log-Normally distributed,

$$r_i \sim \text{LogNormal}(\mu_r, \sigma_r), \quad \text{and} \quad K_i \sim \text{LogNormal}(\mu_K, \sigma_K). \quad (14)$$

For ease of interpretation, Eq (14) is re-parameterised such that $\mathbb{E}[r_i] = \mu_r$, $\text{Var}[r_i] = \sigma_r^2$, $\mathbb{E}[K_i] = \mu_K$, and $\text{Var}[K_i] = \sigma_K^2$. For inference, this leads to a vector of hyper-parameters of interest $\phi = (\mu_r, \sigma_r, \mu_K, \sigma_K)$.

Synthetic data scenarios. Using the particle-cell interaction model (See Example 1: Particle-cell interaction model) and the parameter distributions (Eq. (14)) for r_i and K_i , we generate *in silico* data based on the experimental data collected by Faria et al., [29]. We consider $M=3$ experimental replicates each with $N=20,000$ cells with simulated flow cytometry snapshots taken at $n=6$ time points that are considered identically prepared technical replicates ($t_1 = 1$ [hr], $t_2 = 2$ [hr], $t_3 = 4$ [hr], $t_4 = 8$ [hr], $t_5 = 16$ [hr], and $t_6 = 24$ [hr]). Within this setting, we generate synthetic flow cytometry data for the following three scenarios:

- Scenario 1: All M experimental replicates are i.i.d. with hyper-parameters $\mu_r = 3.86125 \times 10^{-7}$, $\sigma_r = 4.56962 \times 10^{-7}$, $\mu_K = 10$, and $\sigma_K = 2$.
- Scenario 2: Cell parameters in experimental replicate 1 are generated with a larger mean association rate of $\mu_r = 9.0 \times 10^{-7}$. Experimental replicates 2 and 3 are identical to Scenario 1.
- Scenario 3: Cell parameters in experimental replicate 1 are generated with a larger mean association rate $\mu_r = 9.0 \times 10^{-7}$ and cell parameters in experimental replicate 2 are generated with a larger mean association rate of $\mu_r = 9.0 \times 10^{-7}$ and a larger mean carrying capacity $\mu_K = 20$. Experimental replicate 3 is identical to Scenario 1.

In all three scenarios we use, $c=1$, $s = 5.31 \times 10^{-5}$, $v = 1.01 \times 10^{-6}$, and $u_0 = 9.95 \times 10^7$. The specific values for these parameters are based on prior work of Murphy et al., [40].

We can consider Scenario 1 as the ideal experimental conditions for pooled sample analysis, whereas Scenarios 2 and 3 represent different ways the pooled assumptions may not hold. We note that the variation we have introduced to these parameters is fairly modest. Fig 3 provides an example of the synthetic flow cytometry data generated for Scenario 3. The larger mean association rate of replicate 1 (Fig 3(a)) leads to a subtle increase in the location of the peak fluorescence at early times, $t \leq 4$ [hr], compared with replicate 3 (Fig 3(c), the basis of Scenario 1), though at later times, $t > 8$ [hr], replicates 1 and 3 come into alignment. Replicate 2 generally is more diffuse than replicates 1 and 3 for all times, with the peak fluorescence approaching a higher steady state due to the higher mean carrying capacity (Fig 3(b)).

The effect of pooling on quantification of heterogeneity. Before presenting details of the pooled and hierarchical analysis, we focus on estimated distributions of the individual cell parameters resulting from this analysis. In particular, we consider the heterogeneity in the particle association rate (Fig 4(a)) and carrying capacity (Fig 4(b)) that results from the analysis of synthetic data Scenario 3 (See Synthetic data scenarios and Fig 3). This scenario captures a reasonable experimental setting for which the pooled i.i.d. assumptions are invalid due to experimental heterogeneity that is a confounding factor for the biological heterogeneity.

The pooled analysis adequately captures the variability in the cell parameter distributions with support that covers the bulk of each individual replicate distribution (Fig 4). However, the hierarchical analysis highlights that the actual biological heterogeneity is more complex with replicate 3 having a lower particle association rate (Fig 4(a)) and replicate 2 having larger carrying capacity (Fig 4(b)). These replicate specific distributions are not estimated independently, but rather, the hierarchical modelling structure allows for information sharing between replicates to provide quantification of experimental heterogeneity.

Pooled sample analysis. We investigate the sensitivity of a pooled sample analysis to the i.i.d. assumptions. Following the pooled sample analysis protocol of Murphy et al., [40], we estimate the hyper-parameters, $\phi = (\mu_r, \sigma_r, \mu_K, \sigma_K)$, for each of the three synthetic data scenarios. These hyper-parameters characterise the heterogeneity (Eq. (14)) in particle association rates and carrying capacities throughout the cell population, given the cell population dynamics

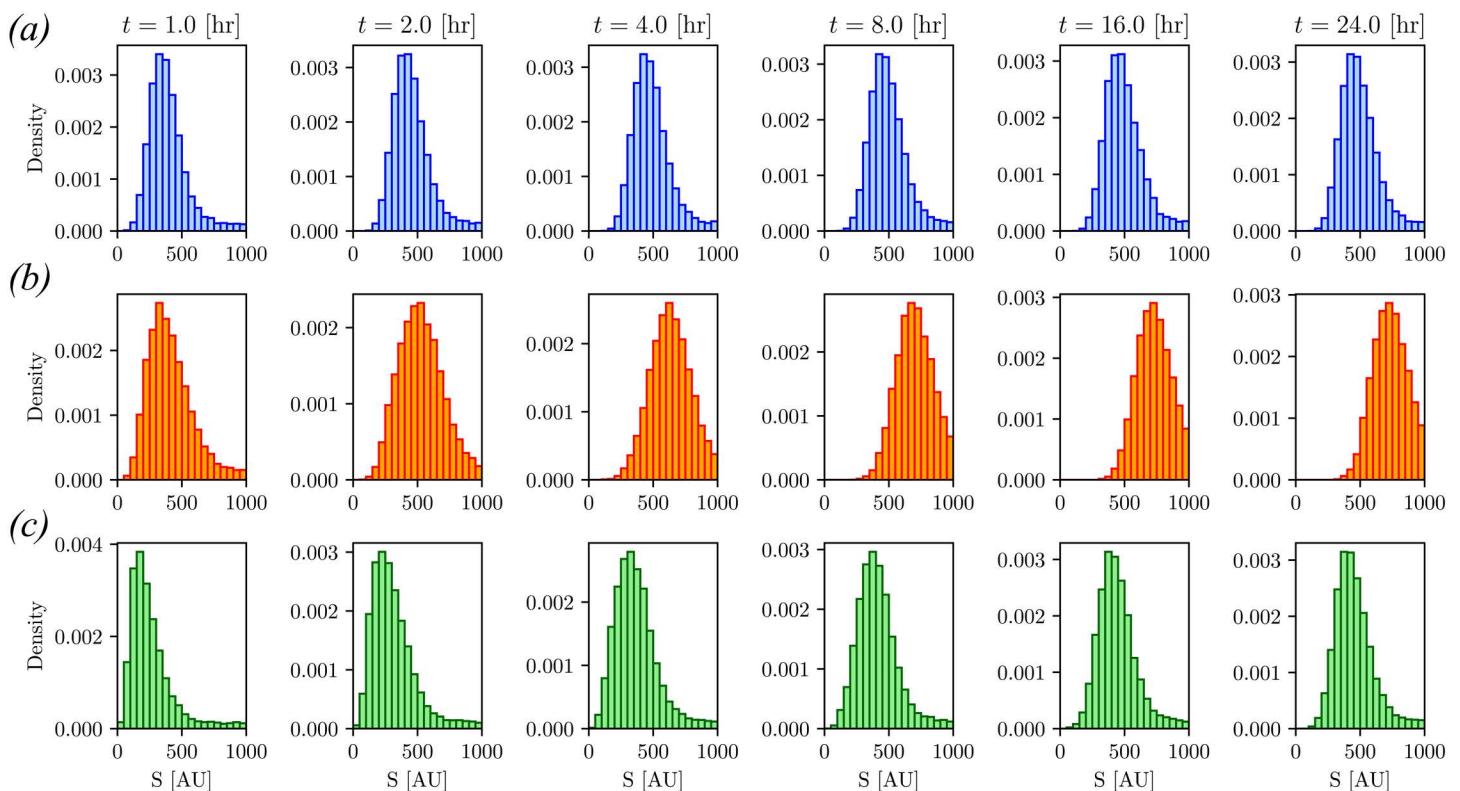


Fig 3. Synthetic flow cytometry population snapshot data example simulating nano-particle cell interaction data. Each histogram represents the simulated fluorescence S [AU] (that is Arbitrary Units) distribution for 20,000 observed cells interacting with nano-particles at one of six observation times ($t_1 = 1$ [hr], $t_2 = 2$ [hr], $t_3 = 4$ [hr], $t_4 = 8$ [hr], $t_5 = 16$ [hr], and $t_6 = 24$ [hr]). Replicate 1 (row (a), blue) is generated with a mean particle association rate of 9.0×10^{-7} and a mean carrying capacity of $\mu_K = 10$. Replicate 2 (row (b), orange) is generated with a mean particle association rate of 3.86125×10^{-7} and a mean carrying capacity of $\mu_K = 20$. Replicate 3 (row (c), green) is generated with a mean particle association rate of 3.86125×10^{-7} and a mean carrying capacity of $\mu_K = 10$. All replicates have the same standard deviation parameters for the association rate and carrying capacity of $\sigma_r = 4.56962 \times 10^{-7}$ and $\sigma_K = 2$, respectively. Here, Replicate 2 (row (b), orange) represents an experimental outlier with higher carrying capacity, resulting in a distinctly larger fluorescence distribution mode at later times than Replicates 1 (row (a), blue) and 3 (row (c), green). Replicate 3 (row (c), green) represents an experimental outlier with a lower association rate, resulting in heavier skewness to the right than Replicates 1 (row (a), blue) and 2 (row (b), orange).

<https://doi.org/10.1371/journal.pcbi.1014379.g003>

under the model of Murphy et al., [40] (See Example 1: Particle-cell interaction model). The Bayesian analysis follows the approach described in the Statistical framework Section, using independent uniform priors, $\mu_r \sim \mathcal{U}(0, 10^{-6})$, $\mu_K \sim \mathcal{U}(0, 50)$, $\sigma_r \sim \mathcal{U}(0, 2 \times 10^{-6})$, and $\sigma_K \sim \mathcal{U}(0, 10)$. These priors are consistent with the analysis of Murphy et al., [40], and cover a range of physically viable values.

The results, as shown in Fig 5, indicate that the estimated hyper-parameters are sensitive to the synthetic data scenario. Firstly, the analysis of Scenario 1 results in posterior distributions that accurately recover the true hyper-parameters of $\mu_r = 3.86125 \times 10^{-7}$ (Fig 5(a)), $\mu_K = 10$ (Fig 5(b)), $\sigma_r = 4.56962 \times 10^{-7}$ (Fig 5(c)), and $\sigma_K = 2$ (Fig 5(d)). This is expected, since Scenario 1 represents our well-specified setting in which each replicate contains identical populations of cells. While this simulation validates the use of the pooled approach when experimental replicates are genuinely the same [39,40], the results for Scenarios 2 and 3 demonstrate substantial sensitivity of the posteriors given deviations in one of the replicates.

In Scenario 2, only one replicate was modified to have a mean association rate of $\mu_r = 9.0 \times 10^{-7}$ resulting in potential outlier in one of the replicates (Fig 3(a)) compared with the unmodified replicates (Fig 3(c)). This outlier replicate strongly

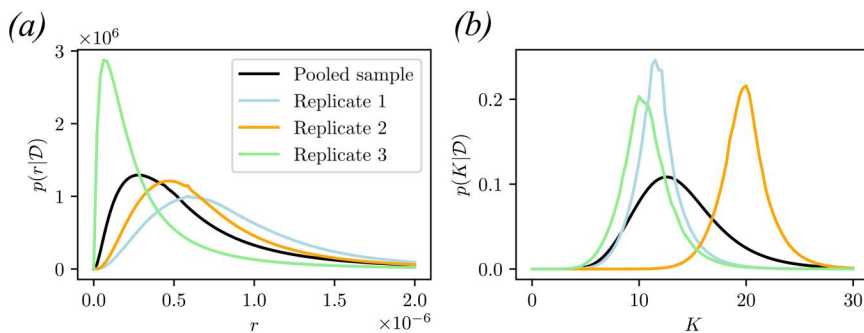


Fig 4. Heterogeneity in posterior cell parameter distributions resulting from pooled and hierarchical analysis of the particle-cell interaction model. Marginal distributions are shown for: (a) the particle association rate, r , and (b) the carrying capacity, K . Data Scenario 3 is used for this analysis (See Synthetic data scenarios and Fig 3). Here, Replicate 2 (orange) represents an experimental outlier with higher mean carrying capacity and Replicate 3 (green) represents an experimental outlier with a lower mean association rate. In this setting the pooled sample (black) does not adequately characterise the combined biological heterogeneity across experimental replicates.

<https://doi.org/10.1371/journal.pcbi.1014379.g004>

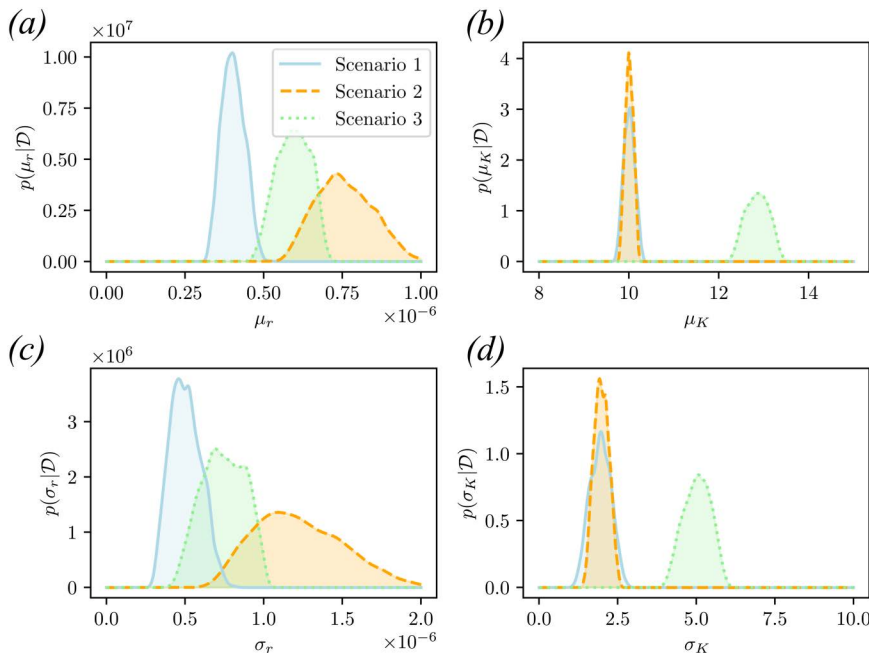


Fig 5. Pooled sample analysis of the nano-engineered particle-cell interaction experiment for the three synthetic data scenarios. As described in Synthetic data scenarios, Scenario 1 (blue) is the i.i.d. case with no outliers, Scenario 2 (orange) has a single outlier replicate with a larger mean association rate, and Scenario 3 (green) has an additional outlier with a larger mean carrying capacity. Marginal posterior distributions for the hyper-parameters: (a) mean association rate, μ_r ; (b) mean carrying capacity, μ_K ; (c) association rate standard deviation, σ_r ; and (d) carrying capacity standard deviation, σ_K .

<https://doi.org/10.1371/journal.pcbi.1014379.g005>

affects the estimates for both the pooled associate rate distributions with inflated means μ_r (Fig 5(a)) and standard deviations σ_r (Fig 5(c)). While the results represent the best pooled sample inference given the outlier replicate, the results may not be meaningful for prediction in biological applications since the inferred heterogeneity does not represent any of the replicates accurately. The results are even more concerning for Scenario 3, where an addition replicate is modified to

be an outlier association rate, $\mu_r = 9.0 \times 10^{-7}$, and in carrying capacity $\mu_K = 20$, since a shift in all four hyper-parameters occurs (Fig 5). This means that the heterogeneity estimates for the pooled population are not representative of any of the three replicate populations. These results motivate the adoption of our extended hierarchical model in experimental settings for which variation is expected between replicates.

Hierarchical analysis. We now demonstrate the advantages of our hierarchical framework (See Statistical framework) as a method for capturing heterogeneity in flow cytometry data when variation between replicates is present. We consider the same synthetic data scenarios (See Synthetic data scenarios) and apply our hierarchical framework to the particle-cell interaction model (See Example 1: Particle-cell interaction model) of Murphy [40]. Specifically, we consider the setting where the i th cell in the j th replicate population has its own associated rate, r_{ij} , and carrying capacity, K_{ij} . Following the pooled sample setting, we assume these parameters $\theta_{ij} = (r_{ij}, K_{ij})$ are log-Normally distributed, however, we deviate from the pooled assumption by allowing each replicate population $j = 1, 2, \dots, M$ to have its own mean association rate, μ_{rj} , and mean carrying capacity, μ_{Kj} ,

$$r_{ij} \sim \text{LogNormal}(\mu_{rj}, \sigma_r), \quad \text{and} \quad K_{ij} \sim \text{LogNormal}(\mu_{Kj}, \sigma_K).$$

Note that the pooled sample model is recovered for the special case where $\mu_{r,1} = \mu_{r,2} = \dots = \mu_{r,M}$ and $\mu_{K,1} = \mu_{K,2} = \dots = \mu_{K,M}$. We further assign a Normal distribution to the replicate population means,

$$\mu_{rj} \sim \mathcal{N}(m_r, s_r), \quad \text{and} \quad \mu_{Kj} \sim \mathcal{N}(m_K, s_K),$$

where m_r (resp. m_K) and s_r (resp. s_K) are the mean and standard deviation of the replicate population association rate (resp. carrying capacity) means. Thus we infer intra-experimental heterogeneity for replicate population j through inference of the hyper-parameters $\phi_j = (\mu_{rj}, \mu_{Kj}, \sigma_r, \sigma_K)$ and inter-experimental heterogeneity through the population level hyper-parameters, $\psi = (m_r, m_K, s_r, s_K)$.

We perform Bayesian inference for the full joint posterior for the two levels of hyper-parameters (Eqs. (8) and (11)). For the population level parameters, we apply independent uniform priors for the means $m_r \sim \mathcal{U}(0, 10^{-6})$, $m_K \sim \mathcal{U}(0, 50)$, and weakly informative half Cauchy priors for the standard deviations $s_r \sim \text{Half-Cauchy}(3 \times 10^{-7})$ and $s_K \sim \text{Half-Cauchy}(0.5)$. Note the shape parameters for the half Cauchy priors are determined based on the guidelines of Gelman et al., [51] with regard to selection of weakly informative priors for standard deviation terms in Bayesian hierarchical models with small numbers of groups.

The results, shown in Fig 6, highlight the value of our hierarchical approach. In Scenario 1 we accurately identify replicates are i.i.d. with $p(\mu_{r,1} | \mathcal{D}) \approx p(\mu_{r,2} | \mathcal{D}) \approx p(\mu_{r,3} | \mathcal{D})$ (Fig 6(a)–6(c)), and $p(\mu_{K,1} | \mathcal{D}) \approx p(\mu_{K,2} | \mathcal{D}) \approx p(\mu_{K,3} | \mathcal{D})$ (Fig 6(e)–6(g)). Furthermore the outlier replicates in Scenario 2 (replicate 1 with larger $\mu_{r,1}$; Fig 6(a)) and Scenario 3 (replicate 1 with larger $\mu_{r,1}$, and replicate 2 with larger $\mu_{K,2}$; Fig 6(a) and 6(f)), are directly captured with only minor inflation in the variance of the posterior distributions for the remaining replicate populations (Fig 6(b) and 6(c) for Scenario 2, and Fig 6(b), 6(c), 6(e) and 6(g) for Scenario 3).

In addition to reliable characterisation of heterogeneity for each replicate population, we also obtain estimates of the inter-experimental heterogeneity that are insensitive to variation in a single replicate. This is evidenced by substantial overlap in the posterior probability densities around the true population level means (Fig 6(i) and 6(j)) and standard deviations (Fig 6(k) and 6(l)). In almost all cases, the true population level hyper-parameter is within the bulk density region of the posterior distribution. The exception, is the population standard deviation parameter for the carrying capacity mean s_K in Scenario 3 (Fig 6(l)), however, the deviation in the posterior distributions for s_K across scenarios is substantially less than that of the equivalent pooled sample analysis (Fig 5(d)). This robustness in population level inferences arises from the between-replicate information sharing inherent to the hierarchical model.

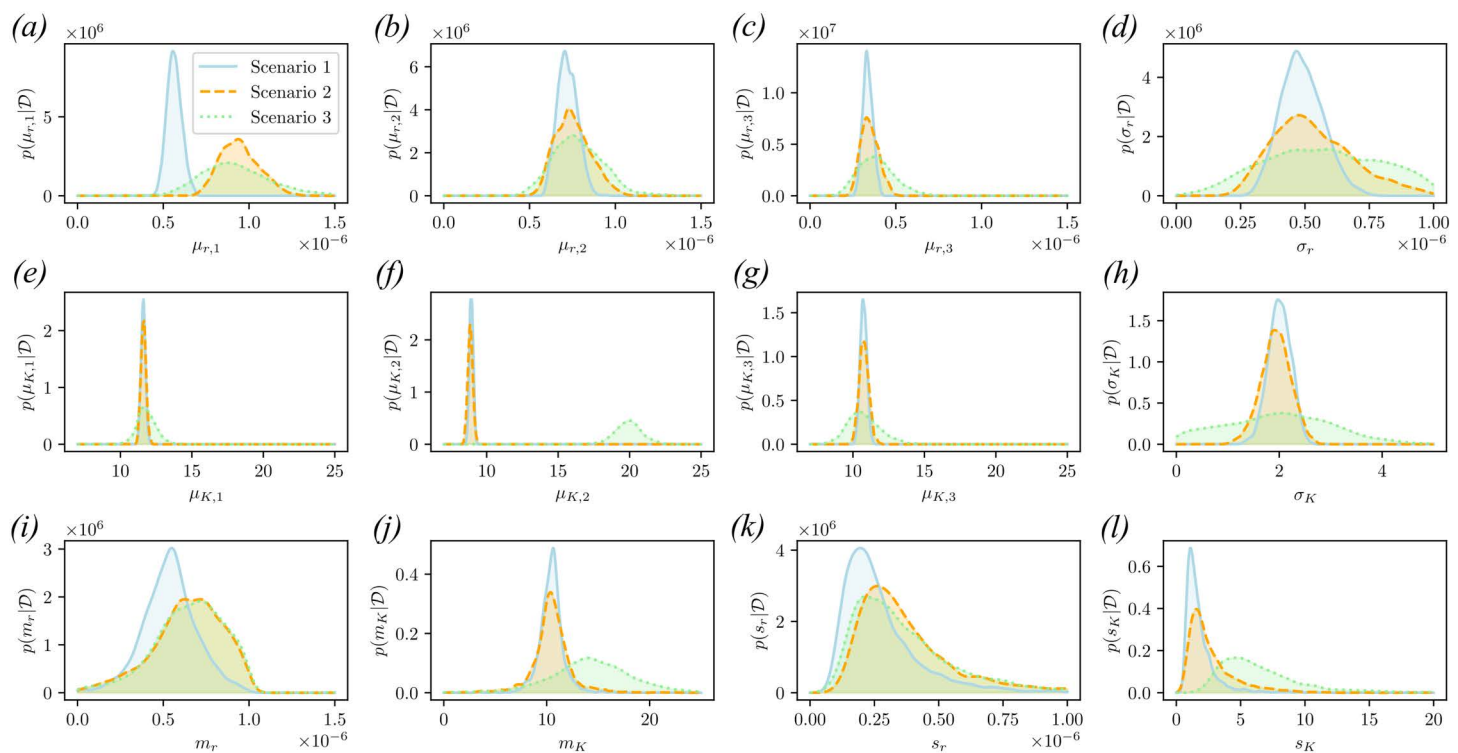


Fig 6. Hierarchical analysis of the nano-engineered particle-cell interaction experiment for the three synthetic data scenarios. As described in Synthetic data scenarios, Scenario 1 (blue) is the i.i.d. case with no outliers, Scenario 2 (orange) has a single outlier replicate with a larger mean association rate, and Scenario 3 (green) has an additional outlier with a larger mean carrying capacity. Marginal posterior distributions are shown for the replicate hyper-parameters: (a)–(c) mean association rates, $\mu_{r,j}$ for replicates $j = 1, 2, 3$; (d) association rate standard deviation, σ_r ; (e)–(g) mean carrying capacity, $\mu_{K,j}$ for replicates $j = 1, 2, 3$; and (h) carrying capacity standard deviation, σ_K . Marginal posterior distribution are shown for the population level hyper-parameters: (i) mean of association rate means, m_r ; (j) mean of carrying capacity means, m_K ; (k) standard deviation of association rate means, s_r ; and (l) standard deviation of carrying capacity means, s_K . Outlier mean association rates ((a), orange and green) and mean carrying capacity ((f), green) are identified, whereas information sharing lead to robust population level heterogeneity estimates (i)–(l).

<https://doi.org/10.1371/journal.pcbi.1014379.g006>

Dual-labelled probe internalisation experiments

To demonstrate the generality of our framework, we perform a similar simulation study based on the study of heterogeneity in cell internalisation processes by Browning et al., [39]. The internalisation process of a cell is modelled according to an ODE system (See Example 2: Internalisation model) with each cell having its own internalisation rate, λ_i , recycling rate, β_i , and initial receptor density R_i . The heterogeneity in these parameters is modelled, in the pooled sample setting, according to

$$\begin{aligned}\lambda_i &\sim \text{ShiftedGamma}(\mu_\lambda, \sigma_\lambda, \omega_\lambda), \\ \beta_i &\sim \text{ShiftedGamma}(\mu_\beta, \sigma_\beta, \omega_\beta), \\ R_{0,i} &\sim \text{ShiftedLogNormal}(\mu_{R_0}, \sigma_{R_0}),\end{aligned}$$

where μ_λ , σ_λ , and ω_λ (resp. μ_β , σ_β , and ω_β) are the mean, standard deviation, and skewness of the internalisation rate (resp. recycling rate). Here, μ_{R_0} , σ_{R_0} are the standard log-Normal parameters and the distribution of $R_{0,i}$ is shifted such that $\mathbb{E}[R_{0,i}] = 1$ (See Browning et al., [39] for further details). For inference, these distribution hyper-parameters are of

interest along with the disassociation rate p , that is, $\phi = (\mu_\lambda, \sigma_\lambda, \omega_\lambda, \mu_\beta, \sigma_\beta, \omega_\beta, \mu_{R_0}, \sigma_{R_0}, p)$. As per Eq. (5), parameters for the observation process, $\eta = (\alpha_1, \alpha_2, \sigma_1, \sigma_2)$, are also inferred, however, we do not focus on these here.

Synthetic data scenarios. For the internalisation model (See Example 2: Internalisation model) we generate synthetic data inspired by the data collected by Browning et al., [39] (See Dual-labelled probe internalisation data). Here, we have $M=3$ replicates, each consisting of quenched/unquenched sample pairs of $N=1,000$ cells each. Simulated flow cytometry snapshots are taken at $n=7$ observation times, $t_1=5$ [min], $t_2=10$ [min], $t_3=20$ [min], $t_4=30$ [min], $t_5=40$ [min], $t_6=120$ [min], and $t_7=180$ [min]. As before (See Synthetic data scenarios), we treat the different snapshots as being generated by identically prepared technical replicates, and thus having the same intra-experimental heterogeneity. We consider two scenarios based:

- Scenario 1: All M replicates are i.i.d., with hyper-parameters $\mu_\lambda = 0.08$, $\sigma_\lambda = 0.01$, $\omega_\lambda = 0.9$, $\mu_\beta = 0.05$, $\sigma_\beta = 0.005$, $\omega_\beta = -1.0$, $\mu_R = 0.05$, and $\sigma_R = 0.01$.
- Scenario 2: Cell parameters in replicate 1 are generated with a smaller mean internalisation rate $\mu_\lambda = 0.02$. Replicates 2 and 3 are identical to Scenario 1.

In both scenarios, $p=0.075$, $\alpha_1 = 8500$, $\alpha_2 = 38$, $\sigma_1 = 5$, and $\sigma_2 = 2$. Parameter values are obtained based on the analysis of Browning et al., [39]. Fig 7 provides an example of the synthetic flow cytometry data generated for Scenario 2.

As with the nano-particle cell interaction example (See Synthetic data scenarios), Scenario 1 represents the setting in which the pooled analysis is expected to be accurate, whereas Scenario 2 includes an outlier replicate (Fig 7(a)) with a smaller internalisation rate. This results in the quenched and unquenched samples being distinguishable throughout all snapshots. By comparison, replicates 2 and 3 lead to almost identical fluorescence distributions for quenched and unquenched at $t=30$ [min].

The effect of pooling on quantification of heterogeneity. In the pooled analysis, we note that the estimated variance of the internalisation rate is inflated to account for the variation across experimental replicates (Fig 8(a)). While the hierarchical approach provides more details with replicate specific parameter distributions, the inflated variance in the pooled approach seems an appropriate global approximation. However, the pooled analysis also inflates the variance of the recycling rate (Fig 8(b)) for which there is no substantial experimental heterogeneity. Conversely the pooled and hierarchical results for the initial receptor density are practically identical (Fig 8(c)). A reasonable explanation for the incorrect inflation of the β distribution is the relationship between λ and β in the model Eq. (4) as receptor recycling is effectively a reversal of internalisation.

Pooled sample analysis. Here we investigate how deviations in the i.i.d. assumption affect quantification of biological heterogeneity, and the impact this has on estimates of global parameters, such as the disassociation probability p . For the pooled sample analysis we follow Browning et al., [39] to estimate the hyper-parameters, $\phi = (\mu_\lambda, \sigma_\lambda, \omega_\lambda, \mu_\beta, \sigma_\beta, \omega_\beta, \mu_{R_0}, \sigma_{R_0}, p)$ that relate to heterogeneity in internalisation rates, recycling rates, and the receptor concentrations. In addition, the purely chemical disassociation probability constant p is also.

The experimental heterogeneity incorporated into data Scenario 2 results in substantial estimated bias and variance inflation for both the mean (Fig 9(a)) and standard deviation (Fig 9(b)) parameters for the internalisation rate. Just as with the particle-cell interaction model (See Nano-engineered particle-cell interaction experiments), the pooled sample analysis results are extremely sensitive to experimental heterogeneity. In addition, inferences of parameters with no biological or experimental heterogeneity, such as the disassociation rate p (Fig 9(c)), are impacted by experimental heterogeneity in the internalisation rate. Other parameter inferences are also inaccurate for Scenario 2, however, we focus here on the internalisation rate and the disassociation rate.

Hierarchical analysis. Expanding the analysis of Browning et al., [39] according to our framework (See Statistical framework) leads to the i th cell in the j th replicate population having its own internalisation rate, λ_{ij} , recycling rate, β_{ij} ,

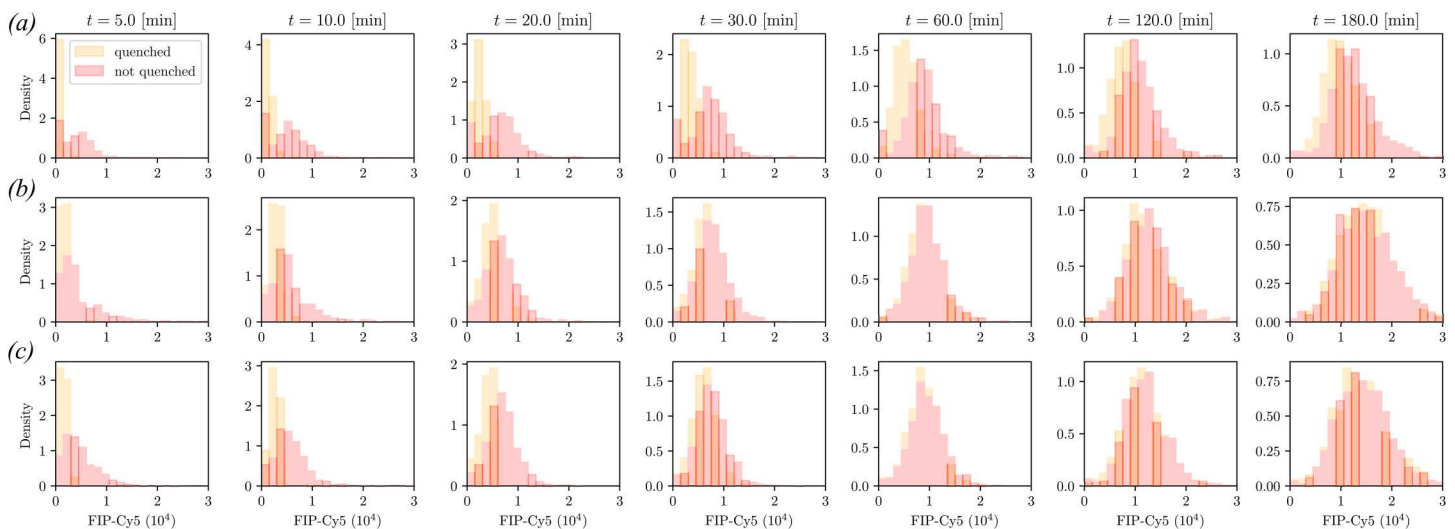


Fig 7. Synthetic flow cytometry population snapshot data example simulating the dual-labelled probe internalisation experiments. Each histogram represent the simulated fluorescence distribution of the quenchable probe (FIP-Cy5) for 1,000 observed cells in the quenched sample (orange) and 1,000 observed cells in the unquenched sample (red) at one of seven observation times ($t_1 = 5$ [min], $t_2 = 10$ [min], $t_3 = 20$ [min], $t_4 = 30$ [min], $t_5 = 60$ [min], $t_6 = 120$ [min], and $t_7 = 180$ [min]). Replicate 1 (row (a)) is generated with a mean internalisation rate of $\mu_\lambda = 0.02$, a mean recycling rate of $\mu_\beta = 0.05$ and a log-normal mean initial receptor concentration of $\mu_R = 0.05$. Replicates 2 and 3 (rows (b) and (c)) are generated with a mean internalisation rate of $\mu_\lambda = 0.08$, a mean recycling rate of $\mu_\beta = 0.05$ and a log-normal mean initial receptor concentration of $\mu_R = 0.05$. All replicates have the same standard deviation parameters, $\sigma_\lambda = 0.01$, $\sigma_\beta = 0.01$, and $\sigma_R = 0.01$. Here Replicate 1 (row (a)) represents an experimental outlier with a lower mean internalisation rate, resulting in a clear differences between quenched (orange) and unquenched (red) fluorescent distributions even at late time. In contrast, both Replicates 2 and 3 (rows (b) and (c)), obtain near identical quenched (orange) and unquenched (red) fluorescent distributions after $t = 60$ [min].

<https://doi.org/10.1371/journal.pcbi.1014379.g007>

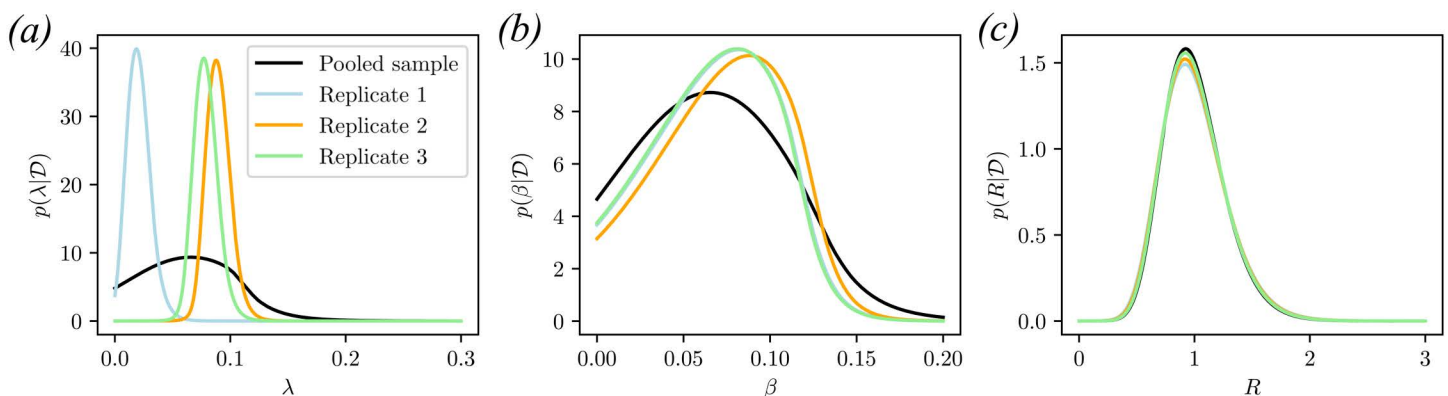


Fig 8. Heterogeneity in posterior cell parameter distributions resulting from pooled and hierarchical analysis of the internalisation model. Marginal distributions are shown for: (a) the internalisation rate, λ , (b) the recycling rate, β , and (c) the initial receptor density, R . Data Scenario 2 is used (See Synthetic data scenarios and Fig 7) which exhibits variation across replicates in the mean internalisation rate. Here, Replicate 1 (blue) represents an experimental outlier with a lower mean internalisation rate (a). In this setting, the pooled sample (black) is not able to capture the biological heterogeneity in internalisation rates (a) or recycling rates (b). The pooled sample is able to quantify biological heterogeneity in the initial receptor density (c) as this parameter is insensitive to the rate parameters.

<https://doi.org/10.1371/journal.pcbi.1014379.g008>

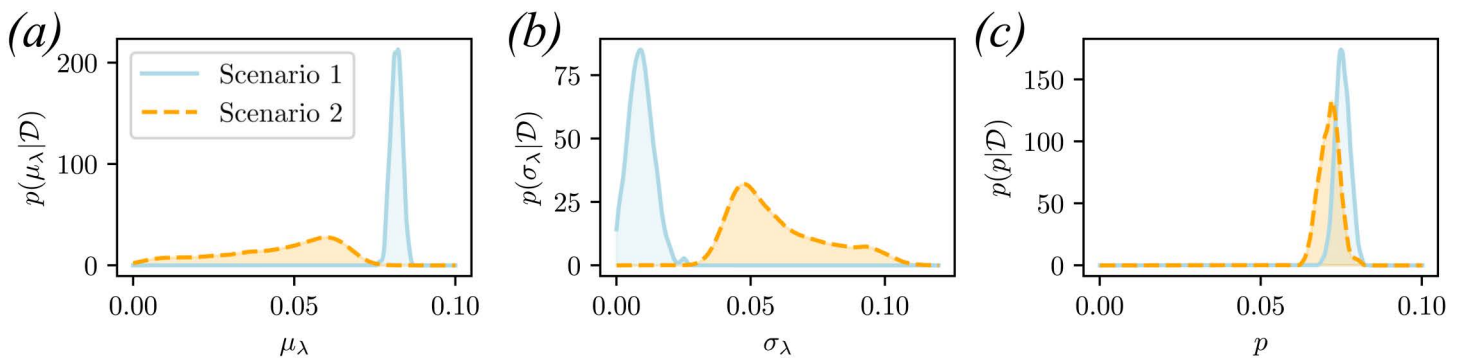


Fig 9. Pooled sample analysis of the dual-label probe internalisation experiment for the two synthetic data scenarios. As described in Synthetic data scenarios, Scenario 1 (blue) is the i.i.d. case with no outliers, and Scenario 2 (orange) has a single outlier replicate with a lower mean internalisation rate. Marginal posterior distributions for the hyper-parameters: **(a)** mean internalisation rate, μ_λ ; **(b)** internalisation rate standard deviation, σ_λ ; and **(d)** the disassociation rate, p .

<https://doi.org/10.1371/journal.pcbi.1014379.g009>

and initial receptor density, R_{ij} . Further we allow each of the replicate populations to have its own mean internalisation rate, $\mu_{\lambda,j}$, mean recycling rate, $\mu_{\beta,j}$, and mean initial receptor density, $\mu_{R,j}$,

$$\begin{aligned}\lambda_{ij} &\sim \text{ShiftedGamma}(\mu_{\lambda,j}, \sigma_\lambda, \omega_\lambda), \\ \beta_{ij} &\sim \text{ShiftedGamma}(\mu_{\beta,j}, \sigma_\beta, \omega_\beta), \\ R_{ij} &\sim \text{ShiftedLogNormal}(\mu_{R,j}, \sigma_R),\end{aligned}$$

and treat the replicate population means as normally distributed,

$$\mu_{\lambda,j} \sim \mathcal{N}(m_\lambda, s_\lambda), \quad \mu_{\beta,j} \sim \mathcal{N}(m_\beta, s_\beta), \quad \text{and} \quad \mu_{R_0,j} \sim \mathcal{N}(m_{R_0}, s_{R_0}).$$

We observe improved results that accurately estimate the mean internalisation rate for each replicate population in both scenarios (Fig 10(a)–10(c)) while quantifying the experimental heterogeneity across replicates (Fig 10(e)–10(f)). Furthermore the disassociation rate is precisely the same in both scenarios (Fig 10(g)). These results highlight that pooled analysis should only be used when there is high certainty in the i.i.d. assumption across replicates.

Discussion

The reliable quantification of heterogeneity in cell populations is crucial to ensure accurate assessment and prediction of treatment efficacy [12–14], especially in the setting of targeted therapeutics [9–11]. Various approaches to modelling and analysis have been developed in the literature to quantify this heterogeneity using cell population snapshots obtained through flow cytometry. However, variability between experimental replicates is rarely considered. In this work, we present a Bayesian hierarchical framework to quantify biological and experimental heterogeneity simultaneously when using mechanistic models of cell dynamics. This is an important development, since previous approaches for quantification of heterogeneity using mechanistic models have only considered pooled samples and neglected experimental heterogeneity [29,30,39,40].

We demonstrated our framework using a two-level hierarchical approach that modelled heterogeneity within experimental replicates and heterogeneity between each replicate. This could naturally be extended to include additional levels that might describe heterogeneity across different labs, instrument configurations, or cell lines. Alternatively, in other experimental settings that support tracking of individual cells [32,52], then the individual cell parameters could also be

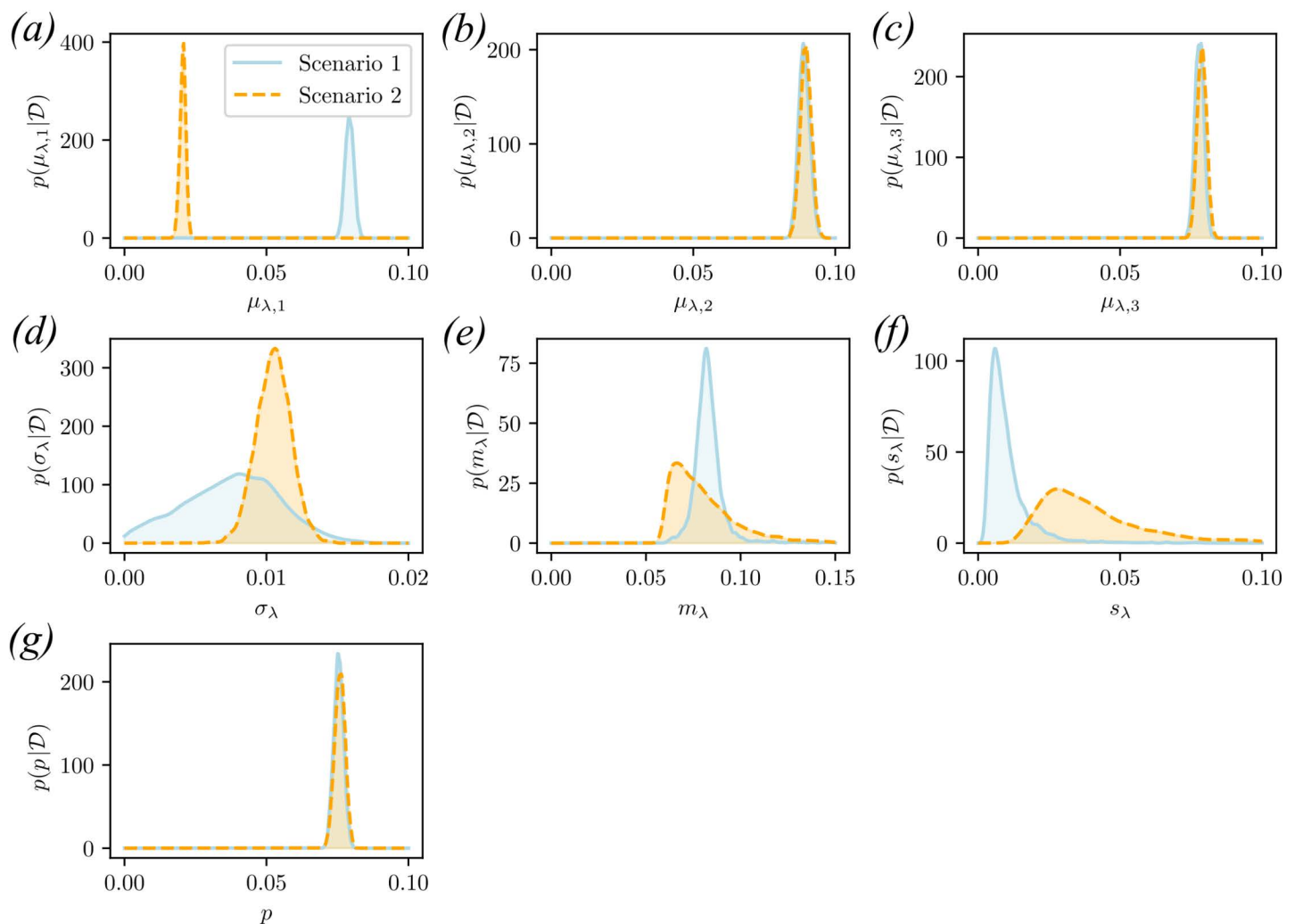


Fig 10. Hierarchical analysis of the dual-labelled probe internalisation experiment for the two synthetic data scenarios. As described in Synthetic data scenarios, Scenario 1 (blue) is the i.i.d. case with no outliers, and Scenario 2 (orange) has a single outlier replicate with a lower mean internalisation rate. Marginal posterior distributions are shown for the replicate hyper-parameters: (a)–(c) mean internalisation rates, $\mu_{\lambda,j}$ for replicates $j = 1, 2, 3$; (d) internalisation rate standard deviation, σ_{λ} . Marginal posterior distributions are shown for the population level hyper-parameters: (e) mean of internalisation rate means, m_{λ} ; (f) standard deviation of internalisation rate means, s_{λ} ; and (g) disassociation rate p . The outlier mean internalisation rate in Scenario 2, Replicate 1 ((a), orange) is identified, while population estimates ((e)–(f)) and homogeneous disassociation rate estimates (g) are robust to this outlier.

<https://doi.org/10.1371/journal.pcbi.1014379.g010>

estimated. This would naturally come with additional computational challenges due to the larger dimensional parameter space, but such challenges are not insurmountable (S1 Appendix).

Our simulation examples assume the scenario in which experimental replicates are repeats of the same experimental conditions. That is, the inter-experimental heterogeneity would be largely to do with variations in the instrumental configuration, operator and variations in uncontrolled aspects of the environment. However, our tools are also entirely applicable to setting in which experimental conditions are also varied, such as different treatments, dosage levels, or cell perturbations [32,52]. As noted in S1 Appendix, such an extension could be introduced with a three-level hierarchical model. As a

result, data integration over multiple experimental conditions could be exploited to reduce certain parameter identifiability problems that are common in single-cell experiments [32,53].

Here, we only consider inter-experimental heterogeneity in the parameter means across replicate populations. Experimental heterogeneity in the parameter variances are mathematically straightforward to introduce. However, some care should be taken to avoid non-identifiability issues arising, especially for small numbers of replicates M [54]. Weakly informative priors, such as the Half-Cauchy prior, are necessary for the variance hyper-parameter of the sub-population parameter means in the low replicate number setting $M=3$ [51], and this would necessarily be more complex with experimental heterogeneity in variances [55]. As a result, we recommend that experimental heterogeneity in the replicate population parameter variances only be considered when the number of replicates, M , is large.

As our method was developed and presented as an extension to the pooled sample approaches [39,40], we utilised similar modelling and analysis decisions. These could be adjusted as appropriate for other applications. For example, we assumed that heterogeneity is driven by variation between individual cells and that stochastic effects are negligible. In rare circumstances analytical solutions are available to account for non-negligible intrinsic noise [56], however, it is often necessary to use stochastic model [49,57]. Since we rely on ABC for parameter inference, our framework is completely applicable to stochastic modelling approaches, including agent-based models or stochastic differential equations (SDEs).

In other settings where intrinsic stochasticity is negligible and the population size is very large, continuum models (in the sense of the observed cell count $N_j \rightarrow \infty$ for all j) could be used [58]. For example population balance models or phenotype-structured partial differential equation models could be plausible in this setting may be appropriate and would enable more complex dynamics to be included, such as cell division and cell death [58–61]. To capture intra-experimental heterogeneity in this case, care must be taken as individual cells are no longer explicitly modelled [62], but rather a continuous cell population density. This renders biologically heterogeneous parameters to be functions of cell density [63]. In effect this renders biological heterogeneity as a random function, rather than N_j i.i.d. samples $\theta_{1,j}, \dots, \theta_{N,j} \sim p(\cdot | \psi_j)$. To characterise biological heterogeneity in this setting, distributions on functions or random measures [64–67] may be needed. This would be an interesting line of future research.

The use of traditional simulation-based inference methods, such as ABC, could be considered a limitation due to the necessity of a tolerance threshold, $\epsilon > 0$, that affects the accuracy of the posterior approximation [68–70]. In addition, targeting a very small ϵ is generally computationally burdensome and modern machine learning approaches are generally considered more efficient in this respect [70–73]. In the non-linear mixed-effects modelling setting, these methods have been extended to achieve more computationally efficient inference through the use of the amortization property to rapidly estimate individual scale likelihoods that can be adapted to obtain population level posteriors [35,36]. However, one desirable property of ABC, particularly in the context of biological modelling, is its robustness to model misspecification or model uncertainty [74–77]. While robust versions of modern machine learning methods continue to be developed [78–84], there is a serious lack of theoretical guarantees on how these methods behave in the misspecified setting. As a result, we still recommend ABC approaches when model misspecification is likely. In situations where ABC is completely computationally infeasible, then the methods of Arruda et al. [35], and Häggström [36] could potentially be extended to our modelling framework. Some care would be needed in dealing with our shared environment component (Eq. (6)), but this would be a promising step forward, especially if ideas from robust simulation-based inference could be incorporated [78–84].

While our approach is developed with flow cytometry experiments in mind [39,40], the method itself is general to any setting with population snapshot data taken from interacting individuals. For example, in the study of marine ecology, hierarchical Bayesian models are frequently used to describe coral reefs that vary across spatial regions [85,86], however, mechanistic descriptions are typically only done on the level of individual reefs [87,88]. Our approach would be well suited to account for spatial heterogeneity in these mechanistic models for more realistic large scale reef forecasts.

Conclusion

We have highlighted that caution must be taken when using pooled sample approaches to avoid potential bias or inaccuracies in quantification of biological heterogeneity. Our approach provides a means of reliable inference when experimental heterogeneity is present. Furthermore, we can quantify the heterogeneity across experiential replicates. This, in turn, could be used to identify potentially problematic replicates and improve experimental protocols. Due to these properties, our method has the potential to greatly enhance the biological insights that can be obtained through flow cytometry and advance developments in quantitative biology. Finally, our approach is generally applicable to other experimental techniques, such as microarrays that can result in drift and change between experimental replicates [89], and to other analysis settings, such as the integration of multiple experimental conditions [32,52]. We provide open-source implementations of our methods on [GitHub](#).

Supporting information

S1 Appendix. Three-level hierarchical extensions.

(PDF)

S2 Appendix. Computational inference.

(PDF)

Acknowledgments

The authors thank Adrienne Jenner and Adriana Zanca for helpful discussions. Computational resources and services used in this work were provided by the eResearch Office, Queensland University of Technology, Brisbane, Australia.

Author contributions

Conceptualization: David J. Warne, Scott A. Sisson, Alexander P. Browning.

Data curation: Stuart T. Johnston, Matthew Faria, Ryan J. Murphy, Alexander P. Browning.

Formal analysis: David J. Warne, Scott A. Sisson.

Funding acquisition: David J. Warne, Stuart T. Johnston, Scott A. Sisson, Alexander P. Browning.

Investigation: David J. Warne, Matthew Faria.

Methodology: David J. Warne, Stuart T. Johnston, Scott A. Sisson, Ryan J. Murphy, Alexander P. Browning.

Project administration: David J. Warne.

Resources: Stuart T. Johnston, Matthew Faria.

Software: David J. Warne, Xiangrun Zhu, Thomas P. Steele, Alexander P. Browning.

Supervision: David J. Warne, Ryan J. Murphy, Alexander P. Browning.

Validation: David J. Warne, Xiangrun Zhu, Thomas P. Steele, Scott A. Sisson, Matthew Faria, Ryan J. Murphy, Alexander P. Browning.

Visualization: David J. Warne, Xiangrun Zhu, Thomas P. Steele.

Writing – original draft: David J. Warne, Xiangrun Zhu.

Writing – review & editing: David J. Warne, Thomas P. Steele, Stuart T. Johnston, Scott A. Sisson, Matthew Faria, Ryan J. Murphy, Alexander P. Browning.

References

- Gough A, Stern AM, Maier J, Lezon T, Shun TY, Chennubhotla C. Biologically relevant heterogeneity: metrics and practical insights. *SLAS Discov*. 2017;22(3):213–37. <https://doi.org/10.1177/2472555216682725>
- Gordon S, Plüddemann A. Tissue macrophages: heterogeneity and functions. *BMC Biol*. 2017;15(1):53. <https://doi.org/10.1186/s12915-017-0392-4> PMID: [28662662](#)
- Wynn TA, Chawla A, Pollard JW. Macrophage biology in development, homeostasis and disease. *Nature*. 2013;496(7446):445–55. <https://doi.org/10.1038/nature12034> PMID: [23619691](#)
- Kwapiszewska K. Physicochemical perspective of biological heterogeneity. *ACS Phys Chem Au*. 2024;4(4):314–21. <https://doi.org/10.1021/acspchemau.3c00079> PMID: [39069985](#)
- Anderson TL, Nande A, Merenstein C, Raynor B, Oommen A, Kelly BJ, et al. Quantifying individual-level heterogeneity in infectiousness and susceptibility through household studies. *Epidemics*. 2023;44:100710. <https://doi.org/10.1016/j.epidem.2023.100710> PMID: [37556994](#)
- Chisholm RH, Lorenzi T, Clairambault J. Cell population heterogeneity and evolution towards drug resistance in cancer: biological and mathematical assessment, theoretical treatment optimisation. *Biochim Biophys Acta*. 2016;1860(11 Pt B):2627–45. <https://doi.org/10.1016/j.bbagen.2016.06.009> PMID: [27339473](#)
- Hilton LK, Scott DW, Morin RD. Biological heterogeneity in diffuse large B-cell lymphoma. *Semin Hematol*. 2023;60(5):267–76. <https://doi.org/10.1053/j.seminhematol.2023.11.006> PMID: [38151380](#)
- Wu VH, Nordin JML, Nguyen S, Joy J, Mampe F, Del Rio Estrada PM, et al. Profound phenotypic and epigenetic heterogeneity of the HIV-1-infected CD4+ T cell reservoir. *Nat Immunol*. 2023;24(2):359–70. <https://doi.org/10.1038/s41590-022-01371-3> PMID: [36536105](#)
- Gewandter JS, McDermott MP, He H, Gao S, Cai X, Farrar JT, et al. Demonstrating heterogeneity of treatment effects among patients: an overlooked but important step toward precision medicine. *Clin Pharmacol Ther*. 2019;106(1):204–10. <https://doi.org/10.1002/cpt.1372> PMID: [30661240](#)
- Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q*. 2004;82(4):661–87. <https://doi.org/10.1111/j.0887-378x.2004.00327.x>
- Pang M, Gabelle A, Saha-Chaudhuri P, Huijbers W, Gafson A, Matthews PM, et al. Precision medicine analysis of heterogeneity in individual-level treatment response to amyloid beta removal in early Alzheimer's disease. *Alzheimers Dement*. 2024;20(2):1102–11. <https://doi.org/10.1002/alz.13431> PMID: [37882364](#)
- Gough AH, Chen N, Shun TY, Lezon TR, Boltz RC, Reese CE, et al. Identifying and quantifying heterogeneity in high content analysis: application of heterogeneity indices to drug discovery. *PLoS One*. 2014;9(7):e102678. <https://doi.org/10.1371/journal.pone.0102678> PMID: [25036749](#)
- Kashyap A, Rapsomaniki MA, Barros V, Fomitcheva-Khartchenko A, Martinelli AL, Rodriguez AF, et al. Quantification of tumor heterogeneity: from data acquisition to metric generation. *Trends Biotechnol*. 2022;40(6):647–76. <https://doi.org/10.1016/j.tibtech.2021.11.006> PMID: [34972597](#)
- Kropivsek K, Kachel P, Goetze S, Wegmann R, Festl Y, Severin Y, et al. Ex vivo drug response heterogeneity reveals personalized therapeutic strategies for patients with multiple myeloma. *Nat Cancer*. 2023;4(5):734–53. <https://doi.org/10.1038/s43018-023-00544-9> PMID: [37081258](#)
- Steininger RJ 3rd, Rajaram S, Girard L, Minna JD, Wu LF, Altschuler SJ. On comparing heterogeneity across biomarkers. *Cytometry A*. 2015;87(6):558–67. <https://doi.org/10.1002/cyto.a.22599> PMID: [25425168](#)
- Adan A, Alizada G, Kiraz Y, Baran Y, Nalbant A. Flow cytometry: basic principles and applications. *Crit Rev Biotechnol*. 2017;37(2):163–76. <https://doi.org/10.3109/07388551.2015.1128876> PMID: [26767547](#)
- Givan AL. Flow cytometry: an introduction. In: *Flow cytometry*. Humana Press; 2010. p. 1–29. https://doi.org/10.1007/978-1-61737-950-5_1
- McKinnon KM. Flow cytometry: an overview. *Curr Protoc Immunol*. 2018;120(1). <https://doi.org/10.1002/cpim.40>
- Altschuler SJ, Wu LF. Cellular heterogeneity: do differences make a difference? *Cell*. 2010;141(4):559–63. <https://doi.org/10.1016/j.cell.2010.04.033>
- Carter B, Zhao K. The epigenetic basis of cellular heterogeneity. *Nat Rev Genet*. 2021;22(4):235–50. <https://doi.org/10.1038/s41576-020-00300-0> PMID: [33244170](#)
- Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science*. 2002;297(5584):1183–6. <https://doi.org/10.1126/science.1070919> PMID: [12183631](#)
- Kell DB, Ryder HM, Kaprelyants AS, Westerhoff HV. Quantifying heterogeneity: flow cytometry of bacterial cultures. *Antonie Van Leeuwenhoek*. 1991;60(3–4):145–58. <https://doi.org/10.1007/BF00430362> PMID: [1725477](#)
- van Merode AEJ, van der Mei HC, Busscher HJ, Waar K, Krom BP. Enterococcus faecalis strains show culture heterogeneity in cell surface charge. *Microbiology (Reading)*. 2006;152(Pt 3):807–14. <https://doi.org/10.1099/mic.0.28460-0> PMID: [16514160](#)
- Banks HT, Kapraun DF, Thompson WC, Peligero C, Argilaguet J, Meyerhans A. A novel statistical analysis and interpretation of flow cytometry data. *J Biol Dyn*. 2013;7(1):96–132. <https://doi.org/10.1080/17513758.2013.812753> PMID: [23826744](#)
- Burel JG, Qian Y, Lindestam Arlehamn C, Weiskopf D, Zapardiel-Gonzalo J, Taplitz R, et al. An integrated workflow to assess technical and biological variability of cell population frequencies in human peripheral blood by flow cytometry. *J Immunol*. 2017;198(4):1748–58. <https://doi.org/10.4049/jimmunol.1601750> PMID: [28069807](#)

26. Galbusera L, Bellement-Theroue G, Urchueguia A, Julou T, van Nimwegen E. Using fluorescence flow cytometry data for single-cell gene expression analysis in bacteria. *PLoS One*. 2020;15(10):e0240233. <https://doi.org/10.1371/journal.pone.0240233> PMID: [33045012](#)
27. Brinkman RR. Improving the rigor and reproducibility of flow cytometry-based clinical research and trials through automated data analysis. *Cytometry A*. 2019;97(2):107–12. <https://doi.org/10.1002/cyto.a.23883>
28. Giesecke C, Feher K, von Volkmann K, Kirsch J, Radbruch A, Kaiser T. Determination of background, signal-to-noise, and dynamic range of a flow cytometer: a novel practical method for instrument characterization and standardization. *Cytometry A*. 2017;91(11):1104–14. <https://doi.org/10.1002/cyto.a.23250> PMID: [28960720](#)
29. Faria M, Noi KF, Dai Q, Björnalm M, Johnston ST, Kempe K, et al. Revisiting cell-particle association in vitro: a quantitative method to compare particle performance. *J Control Release*. 2019;307:355–67. <https://doi.org/10.1016/j.jconrel.2019.06.027> PMID: [31247281](#)
30. Johnston ST, Faria M, Crampin EJ. Isolating the sources of heterogeneity in nano-engineered particle-cell interactions. *J R Soc Interface*. 2020;17(166):20200221. <https://doi.org/10.1098/rsif.2020.0221> PMID: [32429827](#)
31. Hahne F, Khodabakhshi AH, Bashashati A, Wong C, Gascoyne RD, Weng AP, et al. Per-channel basis normalization methods for flow cytometry data. *Cytometry A*. 2009;77A(2):121–31. <https://doi.org/10.1002/cyto.a.20823>
32. Fröhlich F, Reiser A, Fink L, Woschée D, Ligon T, Theis FJ, et al. Multi-experiment nonlinear mixed effect modeling of single-cell translation kinetics after transfection. *NPJ Syst Biol Appl*. 2018;5:1. <https://doi.org/10.1038/s41540-018-0079-7> PMID: [30564456](#)
33. Pyne S, Hu X, Wang K, Rossin E, Lin T-I, Maier LM, et al. Automated high-dimensional flow cytometric data analysis. *Proc Natl Acad Sci U S A*. 2009;106(21):8519–24. <https://doi.org/10.1073/pnas.0903028106> PMID: [19443687](#)
34. Johnsson K, Wallin J, Fontes M. BayesFlow: latent modeling of flow cytometry cell populations. *BMC Bioinform*. 2016;17:25. <https://doi.org/10.1186/s12859-015-0862-z> PMID: [26755197](#)
35. Arruda J, Schälte Y, Peiter C, Teplytska O, Jaehde U, Hasenauer J. An amortized approach to non-linear mixed-effects modeling based on neural posterior estimation. In: Salakhutdinov R, Kolter Z, Heller K, Weller A, Oliver N, Scarlett J, et al., editors. *Proceedings of the 41st International Conference on Machine Learning*. vol. 235 of *Proceedings of Machine Learning Research*. PMLR; 2024. p. 1865–901. Available from: <https://proceedings.mlr.press/v235/arruda24a.html>
36. Häggström H, Persson S, Cvijovic M, Picchini U. Simulation-based inference for stochastic nonlinear mixed-effects models with applications in systems biology. *Stat Comput*. 2026;36(3):99. <https://doi.org/10.1007/s11222-026-10850-8>
37. Hejblum BP, Alkhasim C, Gottardo R, Caron F, Thiébaud R. Sequential Dirichlet process mixtures of multivariate skew t -distributions for model-based clustering of flow cytometry data. *Ann Appl Stat*. 2019;13(1). <https://doi.org/10.1214/18-aos1209>
38. Lee SX, McLachlan GJ, Pyne S. Automated gating and dimension reduction of high-dimensional cytometry data. Springer International Publishing; 2021. p. 281–94. https://doi.org/10.1007/978-3-030-57204-4_16
39. Browning AP, Ansari N, Drovandi C, Johnston APR, Simpson MJ, Jenner AL. Identifying cell-to-cell variability in internalization using flow cytometry. *J R Soc Interface*. 2022;19(190):20220019. <https://doi.org/10.1098/rsif.2022.0019> PMID: [35611619](#)
40. Murphy RJ, Faria M, Osborne JM, Johnston ST. Quantifying biological heterogeneity in nano-engineered particle-cell interaction experiments. *J R Soc Interface*. 2025;22(230):20250206. <https://doi.org/10.1098/rsif.2025.0206> PMID: [40957563](#)
41. FitzGerald LI, Johnston APR. It's what's on the inside that counts: techniques for investigating the uptake and recycling of nanoparticles and proteins in cells. *J Colloid Interface Sci*. 2021;587:64–78. <https://doi.org/10.1016/j.jcis.2020.11.076> PMID: [33370664](#)
42. Selby LI, Aurelio L, Yuen D, Graham B, Johnston APR. Quantifying cellular internalization with a fluorescent click sensor. *ACS Sens*. 2018;3(6):1182–9. <https://doi.org/10.1021/acssensors.8b00219> PMID: [29676153](#)
43. Dumont C, Czuba E, Chen M, Villadangos JA, Johnston APR, Mintern JD. DNA-based probes for flow cytometry analysis of endocytosis and recycling. *Traffic*. 2017;18(4):242–9. <https://doi.org/10.1111/tra.12466> PMID: [28019081](#)
44. Liu H, Johnston APR. A programmable sensor to probe the internalization of proteins and nanoparticles in live cells. *Angew Chem Int Ed Engl*. 2013;52(22):5744–8. <https://doi.org/10.1002/anie.201301243> PMID: [23606601](#)
45. Hasenauer J, Waldherr S, Doszccak M, Radde N, Scheurich P, Allgöwer F. Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinform*. 2011;12:125. <https://doi.org/10.1186/1471-2105-12-125> PMID: [21527025](#)
46. Jin W, Shah ET, Penington CJ, McCue SW, Chopin LK, Simpson MJ. Reproducibility of scratch assays is affected by the initial degree of confluence: experiments, modelling and model selection. *J Theor Biol*. 2016;390:136–45. <https://doi.org/10.1016/j.jtbi.2015.10.040> PMID: [26646767](#)
47. Sisson SA, Fan Y, Beaumont M. Approximate Bayesian computation. Taylor Francis Group; 2018.
48. Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. Approximate Bayesian computation. *PLoS Comput Biol*. 2013;9(1):e1002803. <https://doi.org/10.1371/journal.pcbi.1002803> PMID: [23341757](#)
49. Warne DJ, Baker RE, Simpson MJ. Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art. *J R Soc Interface*. 2019;16(151):20180943. <https://doi.org/10.1098/rsif.2018.0943> PMID: [30958205](#)
50. Drovandi CC, Pettitt AN. Estimation of parameters for macroparasite population evolution using approximate bayesian computation. *Biometrics*. 2011;67(1):225–33. <https://doi.org/10.1111/j.1541-0420.2010.01410.x> PMID: [20345496](#)
51. Gelman A. Bayesian data analysis. 3rd ed. Chapman Hall/CRC Texts in Statistical Science. CRC Press; 2013.

52. Kitanovski S, Conroy S, Sonneck J, Dorsch M, Urban S, Chen J. Uncertainty-aware quantitative analysis of high-throughput live cell migration data. 2025. <https://doi.org/10.1101/2025.06.12.659342>
53. Browning AP, Flegg JA, Murphy RJ. A cautionary tale of model misspecification and identifiability. *Bull Math Biol.* 2025;88(1):5. <https://doi.org/10.1007/s11538-025-01573-4> PMID: 41413679
54. Ogle K, Barber JJ. Ensuring identifiability in hierarchical mixed effects Bayesian models. *Ecol Appl.* 2020;30(7):e02159. <https://doi.org/10.1002/eap.2159> PMID: 32365250
55. Chung Y, Rabe-Hesketh S, Dorie V, Gelman A, Liu J. A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika.* 2013;78(4):685–709. <https://doi.org/10.1007/s11336-013-9328-2>
56. Dowling CV, Ceva PM, Faria M, Johnston ST. On predicting heterogeneity in nanoparticle dosage. *Math Biosci.* 2022;354:108928. <https://doi.org/10.1016/j.mbs.2022.108928> PMID: 36334785
57. Wilkinson DJ. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat Rev Genet.* 2009;10(2):122–33. <https://doi.org/10.1038/nrg2509> PMID: 19139763
58. Ramkrishna D, Singh MR. Population balance modeling: current status and future prospects. *Annu Rev Chem Biomol Eng.* 2014;5:123–46. <https://doi.org/10.1146/annurev-chembioeng-060713-040241> PMID: 24606333
59. Fernandez-de-Cossio-Diaz J, Mulet R. Maximum entropy and population heterogeneity in continuous cell cultures. *PLoS Comput Biol.* 2019;15(2):e1006823. <https://doi.org/10.1371/journal.pcbi.1006823> PMID: 30811392
60. Guilbeteau J, Jain P, Jolly MK, Pouchol C, Pouradier Duteil N. An integrative phenotype-structured partial differential equation model for the population dynamics of epithelial-mesenchymal transition. *NPJ Syst Biol Appl.* 2025;11(1):24. <https://doi.org/10.1038/s41540-025-00502-4> PMID: 40050291
61. Villa C, Maini PK, Browning AP, Jenner AL, Hamis S, Cassidy T. Reducing phenotype-structured partial differential equations models of cancer evolution to systems of ordinary differential equations: a generalised moment dynamics approach. *J Math Biol.* 2025;91(2):22. <https://doi.org/10.1007/s00285-025-02246-5> PMID: 40719870
62. Fuentes-Gari M, Misener R, García-Munzer D, Velliou E, Georgiadis MC, Kostoglou M, et al. A mathematical model of subpopulation kinetics for the deconvolution of leukaemia heterogeneity. *J R Soc Interface.* 2015;12(108):20150276. <https://doi.org/10.1098/rsif.2015.0276> PMID: 26040591
63. Brown A, Stoukides DM, Tzanakakis ES. Quantifying human pluripotent stem cell attributes with population balance modeling. *Stem Cell Res Ther.* 2025;16(1):563. <https://doi.org/10.1186/s13287-025-04668-w> PMID: 41088382
64. Li D, Jones A, Banerjee S, Engelhardt B. Bayesian multi-group gaussian process models for heterogeneous group-structured data. *J Mach Learn Res.* 2025;26:30. PMID: 41019101
65. Savitsky T, Vannucci M, Sha N. Variable selection for nonparametric gaussian process priors: models and computational strategies. *Stat Sci.* 2011;26(1):130–49. <https://doi.org/10.1214/11-STS354> PMID: 24089585
66. Papaspiliopoulos O, Roberts GO. Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika.* 2008;95(1):169–86. <https://doi.org/10.1093/biomet/asn086>
67. Quintana FA, Müller P, Jara A, MacEachern SN. The dependent dirichlet process and related models. *Stat Sci.* 2022;37(1). <https://doi.org/10.1214/20-sts819>
68. Barber S, Voss J, Webster M. The rate of convergence for approximate Bayesian computation. *Electron J Statist.* 2015;9(1). <https://doi.org/10.1214/15-ejs988>
69. Cranmer K, Brehmer J, Louppe G. The frontier of simulation-based inference. *Proc Natl Acad Sci U S A.* 2020;117(48):30055–62. <https://doi.org/10.1073/pnas.1912789117> PMID: 32471948
70. Papamakarios G, Sterratt D, Murray I. Sequential neural likelihood: fast likelihood-free inference with autoregressive flows. In: Chaudhuri K, Sugiyama M, editors. *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics.* vol. 89 of *Proceedings of Machine Learning Research*. PMLR; 2019. p. 837–48. Available from: <https://proceedings.mlr.press/v89/papamakarios19a.html>
71. Frazier DT, Kelly R, Drovandi C, Warne DJ. The statistical accuracy of neural posterior and likelihood estimation. *arXiv:2411.12068* [Preprint]. 2024. <http://arxiv.org/abs/2411.12068> <https://doi.org/10.48550/ARXIV.2411.12068>
72. Papamakarios G, Nalisnick E, Rezende DJ, Mohamed S, Lakshminarayanan B. Normalizing flows for probabilistic modeling and inference. *J Mach Learn Res.* 2021;22(57):1–64. Available from: <http://jmlr.org/papers/v22/19-1028.html>
73. Lueckmann JM, Boelts J, Greenberg D, Goncalves P, Macke J. Benchmarking simulation-based inference. In: Banerjee A, Fukumizu K, editors. *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics.* vol. 130 of *Proceedings of Machine Learning Research*. PMLR; 2021. p. 343–51. Available from: <https://proceedings.mlr.press/v130/lueckmann21a.html>
74. Frazier DT, Robert CP, Rousseau J. Model misspecification in approximate bayesian computation: consequences and diagnostics. *J R Stat Soc Ser B Stat Methodol.* 2020;82(2):421–44. <https://doi.org/10.1111/rssb.12356>
75. Schälte Y, Hasenauer J. Efficient exact inference for dynamical systems with noisy measurements using sequential approximate Bayesian computation. *Bioinformatics.* 2020;36(Suppl_1):i551–9. <https://doi.org/10.1093/bioinformatics/btaa397> PMID: 32657404
76. Wilkinson RD. Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. *Stat Appl Genet Mol Biol.* 2013;12(2):129–41. <https://doi.org/10.1515/sagmb-2013-0010> PMID: 23652634

77. Kelly RP, Warne DJ, Frazier DT, Nott DJ, Gutmann MU, Drovandi C. Simulation-based Bayesian inference under model misspecification. arXiv:2503.12315 [Preprint]. 2025. <http://arxiv.org/abs/2503.12315> <https://doi.org/10.48550/ARXIV.2503.12315>.
78. Huang D, Bharti A, Souza A, Acerbi L, Kaski S. Learning robust statistics for simulation-based inference under model misspecification. In: Oh A, Naumann T, Globerson A, Saenko K, Hardt M, Levine S, editors. Advances in neural information processing systems. vol. 36. Curran Associates, Inc.; 2023. p. 7289–310. Available from: https://proceedings.neurips.cc/paper_files/paper/2023/file/16c5b4102a6b6eb061e502ce6736ad8a-Paper-Conference.pdf
79. Kelly RP, Nott DJ, Frazier DT, Warne DJ, Drovandi C. Misspecification-robust sequential neural likelihood for simulation-based inference. Trans Mach Learn Res. 2024. Available from: <https://openreview.net/forum?id=tbOYJwXhcY>
80. Kelly RP, Frazier DT, Warne DJ, Drovandi CC. Preconditioned robust neural posterior estimation for misspecified simulators. arXiv:2602.18004 [Preprint]. 2026. <http://arxiv.org/abs/2602.18004> <https://doi.org/10.48550/ARXIV.2602.18004>.
81. Mishra A, Habermann D, Schmitt M, Radev ST, Bürkner PC. Robust amortized bayesian inference with self-consistency losses on unlabeled data. arXiv:2501.13483 [Preprint]. 2025. <http://arxiv.org/abs/2501.13483> <https://doi.org/10.48550/ARXIV.2501.13483>
82. Ward D, Cannon P, Beaumont M, Fasiolo M, Schmon S. Robust neural posterior estimation and statistical model criticism. In: Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A, editors. Advances in neural information processing systems. vol. 35. Curran Associates, Inc.; 2022. p. 33845–59. Available from: https://proceedings.neurips.cc/paper_files/paper/2022/file/db0eac6747e3631eb91095cd76065611-Paper-Conference.pdf
83. Wu Y, Radev ST, Tuerlinckx F. Testing and improving the robustness of amortized bayesian inference for cognitive models. Psychol Methods. 2026;31(3):476–500. <https://doi.org/10.1037/met0000814> PMID: 41587193
84. Yuyan W, Evans M, Nott DJ. Robust Bayesian methods using amortized simulation-based inference. arXiv:2504.09475 [Preprint]. 2025. <http://arxiv.org/abs/2504.09475> <https://doi.org/10.48550/ARXIV.2504.09475>
85. Mellin C, Matthews S, Anthony KRN, Brown SC, Caley MJ, Johns KA, et al. Spatial resilience of the Great Barrier Reef under cumulative disturbance impacts. Glob Chang Biol. 2019;25(7):2431–45. <https://doi.org/10.1111/gcb.14625> PMID: 30900790
86. Stuart-Smith RD, Brown CJ, Ceccarelli DM, Edgar GJ. Ecosystem restructuring along the Great Barrier Reef following mass coral bleaching. Nature. 2018;560(7716):92–6. <https://doi.org/10.1038/s41586-018-0359-9> PMID: 30046108
87. Thompson A, Martin K, Logan M. Development of the coral index, a summary of coral reef resilience as a guide for management. J Environ Manage. 2020;271:111038. <https://doi.org/10.1016/j.jenvman.2020.111038> PMID: 32778318
88. Warne DJ, Crossman K, Heron GEM, Sharp JA, Jin W, Wu PP-Y, et al. Mathematical modelling and uncertainty quantification for analysis of biphasic coral reef recovery patterns. Bull Math Biol. 2025;87(9):129. <https://doi.org/10.1007/s11538-025-01512-3> PMID: 40794296
89. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 2010;11(10):733–9. <https://doi.org/10.1038/nrg2825> PMID: 20838408