

# Optimal quantification of contact inhibition in cell populations

David J. Warne<sup>1</sup>, Ruth E. Baker<sup>2</sup>, Matthew J. Simpson<sup>1\*</sup>

<sup>1</sup>School of Mathematical Sciences, Queensland University of Technology,  
Brisbane, Queensland 4001, Australia

<sup>2</sup>Mathematical Institute, University of Oxford,  
Oxford, OX2 6GG, United Kingdom

\*To whom correspondence should be addressed; E-mail: matthew.simpson@qut.edu.au.

September 18, 2017

## Abstract

Contact inhibition refers to a reduction in the rate of cell migration and/or cell proliferation in regions of high cell density. Under normal conditions contact inhibition is associated with the proper functioning tissues, whereas abnormal regulation of contact inhibition is associated with pathological conditions, such as tumor spreading. Unfortunately, standard mathematical modeling practices mask the importance of parameters that control contact inhibition through scaling arguments. Furthermore, standard experimental protocols are insufficient to quantify the effects of contact inhibition because they focus on data describing early time, low-density dynamics only. Here we use the logistic growth equation as a caricature model of contact inhibition to make recommendations as to how to best mitigate these issues. Taking a Bayesian approach we quantify the trade-off between different features of experimental design and estimates of parameter uncertainty so that we can re-formulate a standard cell proliferation assay to provide estimates of both the low-density intrinsic growth rate,  $\lambda$ , and the carrying capacity density,  $K$ , which is a measure of contact inhibition.

Contact inhibition is the tendency of cells to become non-migratory and/or non-proliferative in regions of high cell density [1]. The phenomena of contact inhibition of migration, involving processes such as adhesion, paralysis and contraction [1], is distinct to contact inhibition of proliferation, driven by cell-cell signaling and adhesion [2, 3]; both phenomena are essential for the regulation of structure and function of multicellular organisms.

Down-regulation of contact inhibition of proliferation enhances tumor spreading [4], while wound healing and tissue regeneration also depend crucially on contact inhibition of proliferation [5]. While contact inhibition of proliferation is ubiquitous in both normal and pathological processes, it is difficult to quantify the impact of such contact inhibition in complex biological systems despite the availability of experimental data. Therefore, mathematical models have an important role in informing our understanding of how contact inhibition of proliferation affects collective cell behavior.

The most fundamental mathematical model describing contact inhibition of cell proliferation is the logistic growth model [6, 7, 8, 9],

$$\frac{dC(t)}{dt} = \underbrace{\lambda C(t)}_{\text{proliferation}} \times \underbrace{\left[1 - \frac{C(t)}{K}\right]}_{\text{contact inhibition}}, \quad (1)$$

where  $C(t) > 0$  is the cell density at time  $t$ ,  $\lambda > 0$  is the proliferation rate, and  $K > 0$  is the carrying capacity density. The carrying capacity density is the density at which contact inhibition decreases the net growth rate to zero. The logistic growth model is used ubiquitously in the study of development, repair and tissue regeneration, including for the modeling of tumor growth [10, 11, 12, 13] and wound healing [6, 14, 15]. The solution of Eq. (1),

$$C(t) = \frac{C(0)K}{[(K - C(0)) \exp(-\lambda t) + C(0)]}, \quad (2)$$

is a sigmoid curve that increases from  $C(t) = C(0)$  to  $C(t) = K$  as  $t \rightarrow \infty$ , provided that  $C(0)/K \ll 1$ .

*In vitro* cell proliferation assays are used routinely to examine mechanisms that control cell proliferation, such as the application of various drugs and other treatments on the rate of cell proliferation [16, 17]. *In vitro* assays are routinely used to inform the development and interpretation of *in vivo* assays describing pathological situations, such as tumor growth [18]. Therefore, improving the design and interpretation of *in vivo* assays will have an indirect influence on the way that we design and interpret *in vivo* assays.

A cell proliferation assay typically involves placing cells, at low density,  $C(0)/K \ll 1$ , onto a two-dimensional surface and re-examining the increased monolayer density at a later time,  $t = T$ . Typical data from a cell proliferation assay are given in Fig. 1 A-B.

To use Eq. (1) to quantitatively inform our understanding of a particular biological system, we must be able to estimate the initial density,  $C(0)$ , and the model parameters,  $\lambda$  and  $K$ . Obtaining an accurate estimate of  $K$  is crucial to understand how contact inhibition controls net the proliferation rate at modest to high densities. However, despite the importance of  $K$ , most theoretical studies work with a non-dimensionalized model by setting  $c(t) = C(t)/K$ . This leads to [6, 7, 14]

$$\frac{dc(t)}{dt} = \lambda c(t) [1 - c(t)],$$

which completely masks the importance of being able to accurately estimate the carrying capacity,  $K$ .

Not only do standard mathematical approaches prevent quantitative assessment of the impact of contact inhibition, in addition, standard experimental protocols for *in vitro* proliferation assays are also insufficient to estimate  $K$ . Very recently, Jin et al., [19] used Eq. (1) to analyze a set of cell proliferation assays performed with a prostate cancer cell line, and concluded that standard experimental data do not lead to robust, biologically relevant estimates of  $K$ . This is consistent with the work of Sarapata and de Pillis who also find that standard *in vitro* experimental data are insufficient to estimate  $K$  using best-fit, non-linear least-squares methods [20]. The fundamental issue, as illustrated in Fig. 1C, is that cell proliferation assays are initiated with a small density and performed for a relatively short duration. This strategy is sufficient for estimating the low-density intrinsic proliferation rate,  $\lambda$ , but completely inadequate for estimating  $K$  which is associated with longer time, higher density data. Given the importance of  $K$ , we are motivated to re-consider the design of proliferation assays so that we can quantitatively estimate both  $\lambda$  and  $K$  from a single experiment.

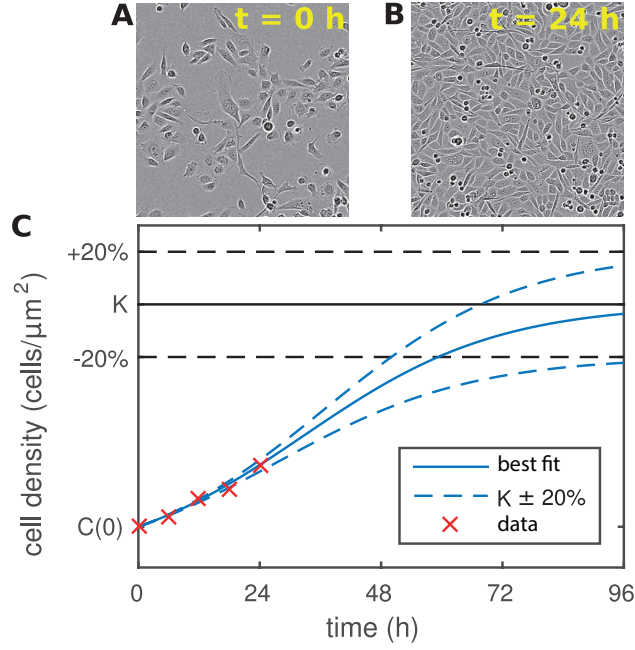


Figure 1: Proliferation assay using a prostate cancer cell line. Images of area  $450 \mu\text{m}^2$  are captured at (A)  $t = 0$  h and (B)  $t = 24$  h. Images are reproduced from Jin et al., (2017), with permission. (C) Logistic growth curve with  $C(0) = 3.1 \times 10^{-4}$  (cells/ $\mu\text{m}^2$ ),  $\lambda = 0.052$  (1/h) and  $K = 2 \times 10^{-3}$  (cells/ $\mu\text{m}^2$ ) (solid black). Additional solutions with  $K \pm 20\%$  (dashed black) also fit the short time experimental data (red crosses).

The typical duration of a proliferation assay is less than 24 h [16], with some assays as short as 4 h [21], and the cell density is recorded once, at the end point of the experiment. This timescale is sufficient to estimate the low-density intrinsic proliferation rate,  $\lambda$ , since the doubling time of the majority of cell lines is approximately 24 hours [6, 7, 13, 15]. However, this standard timescale is far too short to robustly quantify contact inhibition effects given the low initial densities that are routinely used. For example, the PC-3 prostate cancer cell line examined by Jin et al. [19] proliferates with  $\lambda \approx 0.05$  /h, but on the timescale of the experiment the observed growth of the population is approximately exponential. This is consistent with Eq. (1) since we have  $C(t) \sim C(0) \exp(\lambda t)$  provided that  $C(0)/K \ll 1$  and  $t$  is sufficiently small. Fig. 1C shows that the early time growth dynamics are effectively independent of  $K$ , confirming that it is impossible to obtain robust estimates of  $K$  using standard data [19, 20].

The aim of this work is to provide guidance about how to overcome these standard experimental and theoretical limitations by taking a Bayesian approach to experimental design [22, 23, 24, 25, 26, 27]. The advantage of taking a Bayesian approach is that we have a platform to quantitatively examine how the uncertainty in our estimate of  $K$  depends on the experimental design: we aim to provide guidance for experimental design that minimizes the uncertainty in our estimate of  $K$ . To achieve this we consider a proliferation assay of duration  $T$ , with  $n$  observations of cell density,  $C_{\text{obs}}^{1:n} = [C_{\text{obs}}(t_1), \dots, C_{\text{obs}}(t_n)]$ , at times  $t_1, \dots, t_n$  with  $0 < t_i \leq T$  for all  $i = [1, \dots, n]$ . These data could represent a single experiment observed at multiple time points,  $t_1 < t_2 < \dots < t_n \leq T$ , or  $n$  identically prepared experiments, each of which is observed once,  $t_1 = t_2 = \dots = t_n = T$ . We assume that cells proliferate according to Eq. (1) with known  $\lambda$  and  $C(0)$ . Furthermore, we assume  $C(0)/K \ll 1$ . Our assumption that  $\lambda$  can be determined is reasonable since  $C(t) \sim C(0) \exp(\lambda t) = C(0)[1 + \lambda t + \mathcal{O}(t^2)]$  for  $C(0)/K \ll 1$ . This means that fitting a simple straight line or exponential curve to typical experimental data

will provide a reasonable estimate of  $\lambda$ . The assumption that  $C(0)$  is known precisely is less realistic. For example, estimates of  $C(0)$  are affected profoundly by fluctuations in the initialization of the experiment as proliferation assays are performed by placing a known number of cells onto a tissue culture plate. However, images of the experiments are obtained over a much smaller spatial scale. This means that the role of stochastic fluctuations can be significant [19].

To make progress we assume that observations are made, and are subject to Gaussian-distributed experimental measurement error with zero mean and variance  $\Sigma^2$ . Under these conditions our knowledge of the carrying capacity,  $K$ , given such observations is represented by the probability density function

$$p(K | C_{\text{obs}}^{1:n}) = A \prod_{i=1}^n \phi(C_{\text{obs}}(t_i); C(t_i), \Sigma^2), \quad (3)$$

where  $A$  is a normalization constant and  $\phi(\cdot; C(t_i), \Sigma^2)$  denotes a Gaussian probability density with mean  $C(t_i)$  and variance  $\Sigma^2$  (Supporting Material). This probability density represents knowledge obtained from the data when no prior assumptions are made on  $K$ . The point of maximum density in Eq. (3) corresponds to the maximum likelihood estimator or best-fit estimate,  $\hat{K}$ . Importantly, Eq. (3) also enables the quantification of uncertainty in this estimate through calculation of the variance,

$$\sigma_n^2 = \int_0^\infty (K - \hat{K})^2 p(K | C_{\text{obs}}^{1:n}) dK. \quad (4)$$

Fig. 2A shows the probability density of  $K$  (Eq. (3)) for several values of  $T$  for the typical assay protocol where only a single observation is made ( $n = 1$ ). Here, our estimates of  $\lambda$ ,  $C(0)$  and  $\Sigma$  are based on reported values [19]. The spread of these curves indicates the degree of uncertainty in any estimate of  $K$ . In particular, note the red line, which indicates the probability density for  $K$  for a measurement taken at the standard duration of  $T = 24$  h. The relatively flat, disperse nature of the profile confirms that standard proliferation assay designs are completely inappropriate to estimate  $K$  since the profile lacks a well-defined maximum. This result provides a formal explanation for the observations of both Sarapata and de Pillis [20] and Jin et al. [19]. In response to this issue, here we provide quantitative guidelines about how the experimental design can be chosen to facilitate accurate quantification of the effects of contact inhibition.

One optimistic assumption in Eq. (3) is the supposition that  $C(0)$  is known precisely. In reality,  $C(0)$  is subject to both measurement errors and systematic errors owing to stochastic fluctuations [19]. We extend our analysis to incorporate uncertainty in the estimate of  $C(0)$  by also assuming  $C(0)$  to be Gaussian-distributed with mean  $\mu_0$  and variance  $\Sigma_0^2$ , that is  $p(C(0)) = \phi(C(0); \mu_0, \Sigma_0^2)$ . In this case Eq. (3) generalizes to (Supporting Material)

$$p(K | C_{\text{obs}}^{1:n}, \mu_0, \Sigma_0^2) = \int_0^\infty p(K | C_{\text{obs}}^{1:n}) p(C(0)) dC(0). \quad (5)$$

The integral in Eq. (5) is intractable, so numerical integration is required to evaluate  $p(K | C_{\text{obs}}^{1:n}, \mu_0, \Sigma_0^2)$ .

Since  $C(t_i) \rightarrow K$  as  $t_i \rightarrow \infty$  for all  $i = 1, 2, \dots, n$ , then, for both Eq. (3) and Eq. (5), we obtain (Supporting Material)

$$\lim_{(t_1, \dots, t_n) \rightarrow (\infty, \dots, \infty)} p(K | C_{\text{obs}}^{1:n}) = \phi\left(K; \frac{1}{n} \sum_{i=1}^n C_{\text{obs}}^i, \frac{\Sigma^2}{n}\right). \quad (6)$$

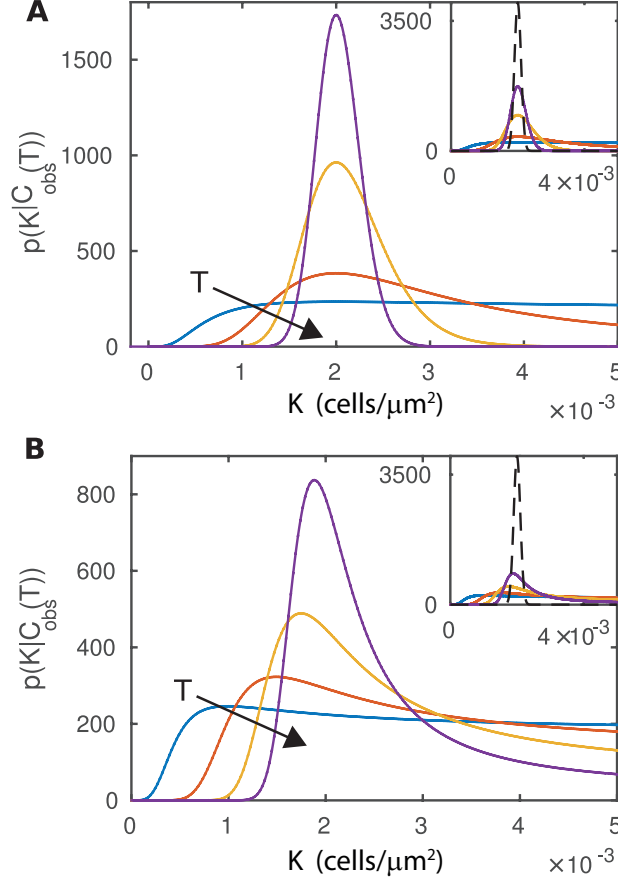


Figure 2: The probability density,  $p(K | C_{\text{obs}}^{1:n})$ , with  $n = 1$  plotted for  $T = 12$  h (black), 24 h (red), 36 h (yellow) and 48 h (purple), where  $\lambda = 5.2 \times 10^{-2}$  (1/h),  $\sigma = 10^{-4}$  (cells/ $\mu\text{m}^2$ ) and the true carrying capacity is  $K = 2 \times 10^{-3}$  (cells/ $\mu\text{m}^2$ ). (A) The initial density,  $C(0)$ , is assumed to be known precisely,  $C(0) = 3.1 \times 10^{-4}$  (cells/ $\mu\text{m}^2$ ). (B) Including uncertainty in the initial condition with  $\mu_0 = 3.1 \times 10^{-4}$  (cells/ $\mu\text{m}^2$ ) and  $\Sigma_0 = 1.02 \times 10^{-4}$  (cells/ $\mu\text{m}^2$ ). The inset panels in both (A) and (B) show the main plot in the context of the limiting density as  $T \rightarrow \infty$  (black dashed line)

Fig. 2A demonstrates that the probability density function for  $K$ , Eq. (3), tightens toward the limiting density, Eq. (6), as  $T$  increases. Note that for the typical assay duration,  $T < 24$  h, the density is approximately uniform. Again, this reiterates the fact that standard experimental designs are inadequate to characterize the effects of contact inhibition, and that different approaches are required. The effect of uncertainty in  $C(0)$  is clear in Fig. 2B. Even with  $T = 48$  h there is a very large region of non-zero, near-constant probability density, indicating very wide confidence intervals for the estimate of  $K$ . Eq. (6) also provides a lower bound on the uncertainty in the estimate of  $K$ ,

$$\sigma_n^2 \geq \frac{\Sigma^2}{n}, \quad (7)$$

for any choice of  $C(0)$  and  $\lambda$ . This result is independent of the treatment of the uncertainty in  $C(0)$ , but requires observations to be made after an infinite duration of time.

Eq. (6) and Eq. (7) tell us two important things about assay design in the study of contact inhibition. First, there is a fundamental lower bound on the uncertainty in our estimate of  $K$

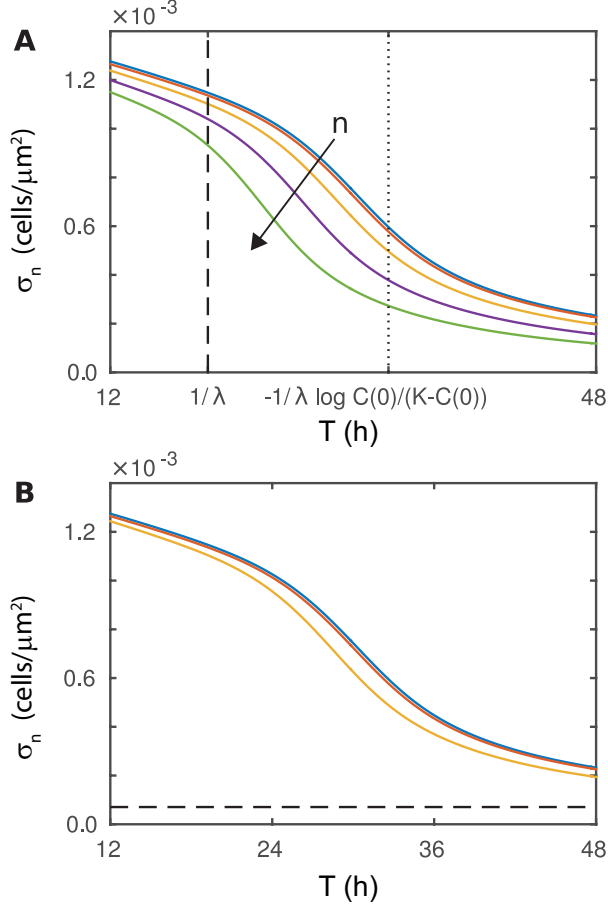


Figure 3: (A) Uncertainty in  $K$  as a function of  $T$  for  $n = 1$  (black),  $n = 2$  (red),  $n = 4$  (yellow),  $n = 8$  (purple) and  $n = 16$  (green); here observations are taken at regular intervals. (B) Effect of observation placement for  $n = 2$  with  $t_2 = T$ ;  $t_1 = T/5$  (black),  $t_1 = T/2$  (red) and  $t_1 = 4T/5$  (yellow). The lower bound on uncertainty,  $\sigma_n$ , for  $n = 2$  (dashed). The uncertainty,  $\sigma_n$ , is calculated using the trapezoid rule with  $10^5$  equally spaced panels over the interval  $0 < K < 5 \times 10^{-3}$  ( $\text{cells}/\mu\text{m}^2$ ).

for a fixed number,  $n$ , of observations of the density. However, increasing the assay duration,  $T$ , always provides more information. Second, increasing the number of observations,  $n$ , always provides more information and, further, it decreases the long time lower bound on the uncertainty in the estimate of  $K$ . Hence, Eq. (7) informs the minimum number of observations required to estimate  $K$  accurately.

Clearly, practical experimental designs require finite  $T$ , and so we require methods to determine  $T$  such that the uncertainty in  $K$  is sufficiently close to the lower bound. We can quantify, for the standard choice of  $n = 1$ , the approximate uncertainty in our estimate of  $K$  (Supporting Material) [28], given by

$$\sigma_1^2 = \frac{C(0)^4(1 - \exp(-\lambda T))^2}{[C(0) - C_{\text{obs}}(T)\exp(-\lambda T)]^4} \Sigma^2. \quad (8)$$

This estimate of the uncertainty in  $K$  is accurate provided  $C_{\text{obs}}(T)/\Sigma^2 \gg 1$ . This expression provides a practical tool to assess the information content of data, but also enables one to estimate whether  $T$  is large enough that the uncertainty in  $K$  is sufficiently close to the lower bound (Supporting Material).



Eq. (7) gives us a method of quantifying the information gained by introducing more observations in the idealized case that  $T$  is sufficiently large. For example, doubling  $n$  will half the uncertainty. However, since practical limitations mean that  $T$  is finite, this result does not always hold. Fig. 3A shows that the effect of doubling  $n$  varies significantly depending on the choice of  $T$ . Here, the  $n$  observations are taken at regular intervals. For example, if  $T < 1/\lambda$ , increasing  $n$  has almost no effect on the uncertainty,  $\sigma_n^2$ . There is a similar negligible effect for  $T > -1/\lambda \log C(0)/(K - C(0))$ , which is the time corresponding to the point of inflection of Eq. (2).

These results highlight several subtle, but immensely important considerations. For example, at sufficiently short times increasing  $n$  has very benefit. Similarly, at sufficiently large times increasing  $T$  has little benefit. The most useful result is that for intermediate times there is more value in increasing  $T$  than increasing  $n$ . Furthermore, if we wish to go beyond the standard experimental design where a single measurement is made at the end of the experiment,  $t = T$ , we might want to quantify the benefit of making a second observation at an earlier time,  $t < T$ , during the same experiment. In this scenario, results in Fig. 3B show that the choice of time at which the second measurement is made can be important. Comparing Fig. 3A with Fig. 3B shows that a poor choice of the time for the second observation might not lead to any change than taking a single observation at  $T$ . However, selecting two well-placed observation times can be as informative as making four equally spaced observations.

We conclude by providing several guidelines for the design of cell proliferation assays:

1. Reducing the uncertainty in  $C(0)$  is crucial, especially if  $T < 48$  h;
2. Eq. (8) should be used with short timescale data to estimate the smallest value of  $T$  that will result in acceptable uncertainty in  $K$ ;
3. On a short timescale, increasing  $T$  is more informative than increasing  $n$ . However, if increasing  $T$  is infeasible, the optimal strategy is to repeat the experiment  $n$  times and make  $n$  observations at time  $T$ . In many situations,  $n$  will need to be large to account for the short timescale. Fig. 3A shows that even a 16-fold increase in  $n$  is still unacceptable for  $T < 12$ .

In this work we highlight certain aspects of assay design that are often neglected and unreported. However, these features are critical if we are to quantify the role of crowding and contact inhibition of proliferation in populations of cells. The guidelines we propose allow us to provide the best estimates of  $\lambda$  and  $K$  using a single experiment, whereas standard experimental designs allow us to confidently estimate  $\lambda$  only. Our results confirm that standard *in vitro* experimental designs are well suited for estimating  $\lambda$ , but poorly suited for estimating  $K$ . A different approach is required to overcome this limitation, and here we provide quantitative guidance about designing *in vitro* proliferation assays that can be used to estimate both  $\lambda$  and  $K$ . The situation is even more complex for *in vivo* assays in which there are more experimental constraints and unknowns. However, improving the way that we design and interpret *in vitro* assays is relevant because these simpler experiments are routinely used in tandem with *in vivo* assays due to the fact that they are cheaper and faster to perform than working in live tissues.

## Author Contributions

DJW, REB and MJS designed the research. DJW performed the research. DJW, REB and MJS contributed analytic tools. DJW, REB and MJS analyzed the data. DJW, REB and MJS wrote the paper.

# ACKNOWLEDGEMENTS

MJS thanks the Australian Research Council (DP170100474). REB thanks the Royal Society for a Royal Society Wolfson Research Merit Award, and the Leverhulme Trust for a Leverhulme Research Fellowship. We thank Michael Stumpf and two anonymous referees for helpful comments.

# APPENDIX

## Posterior probability density functions

In Bayesian statistics, knowledge of unobserved model parameters,  $\theta$ , given the results of some experiment that results in data,  $\mathcal{D}$ , is represented through a probability density function (PDF) [22]. This PDF,  $p(\theta | \mathcal{D})$ , is called the *posterior* and can be interpreted as “the probability density of  $\theta$  given observation  $\mathcal{D}$ ”. The posterior is derived through Bayes’ Theorem,

$$p(\theta | \mathcal{D}) = \frac{\mathcal{L}(\theta; \mathcal{D})p(\theta)}{p(\mathcal{D})}, \quad (\text{A.1})$$

where the *prior* PDF,  $p(\theta)$ , represents knowledge before the experiment, the *likelihood* function,  $\mathcal{L}(\theta; \mathcal{D})$ , determines the probability density of the experimental results,  $\mathcal{D}$ , for a given set of parameter values and the *evidence*,  $p(\mathcal{D})$ , is the likelihood taken across all parameter values. In effect,  $\mathcal{L}(\theta; \mathcal{D})$  encodes assumptions of the model,  $p(\theta)$  encodes the assumptions on the parameters and  $p(\mathcal{D})$  acts as a normalization constant. Our task is to derive the posterior PDF for the carrying capacity density,  $K$ , given noisy observations of an assumed logistic growth curve.

First we derive the likelihood. We represent the process of taking a cell density measurement,  $C_{\text{obs}}(t)$ , at time,  $t$ , as a Gaussian random variable with mean around the true cell density,  $C(t)$  (Eq. (2)), and variance  $\sigma^2$ . Therefore, the probability density of a single observation is a Gaussian PDF,

$$\begin{aligned} \phi(C_{\text{obs}}(t); C(t), \sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-(C_{\text{obs}}(t)-C(t))^2/(2\sigma^2)} \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(C_{\text{obs}}(t) - \frac{C(0)K}{(K-C(0))e^{-\lambda t} + C(0)}\right)^2/(2\sigma^2)}. \end{aligned} \quad (\text{A.2})$$

Assuming  $C(0)$  and  $\lambda$  are known, the exact logistic growth curve is fully determined for any proposed value of  $K$ . As a result, the observations made at time  $t_1$  and  $t_2$ ,  $C_{\text{obs}}(t_1)$  and  $C_{\text{obs}}(t_2)$ , are independent given this value of  $K$ . Therefore the likelihood function for  $n$  observations,  $C_{\text{obs}}^{1:n} = [C_{\text{obs}}(t_1), \dots, C_{\text{obs}}(t_n)]$ , is given by

$$\mathcal{L}(K; C_{\text{obs}}^{1:n}) = \prod_{i=1}^n \phi(C_{\text{obs}}(t_i); C(t_i), \sigma^2), \quad (\text{A.3})$$

where  $\phi(C_{\text{obs}}(t_i); C(t_i), \sigma^2)$  is given by Eq. (A.2).

For the prior, we assume  $0 \leq K \leq K_{\text{max}}$  with equal probability for some  $K_{\text{max}} < \infty$ , that is  $K$  is uniformly distributed on the interval  $(0, K_{\text{max}})$ . The PDF is

$$p(K) = \begin{cases} \frac{1}{K_{\text{max}}}, & \text{if } K \in [0, K_{\text{max}}], \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A.4})$$



This uniform prior distribution imposes minimal assumptions, as we impose no prior knowledge about the value of  $K$  outside of some upper bound that could be arbitrarily large. For uninformative priors, the modes of the posterior correspond to the maximum likelihood estimator (MLE).

The evidence acts as a normalizing constant to ensure the product of Eq. (A.3) and Eq. (A.4) is a true PDF. Therefore, we have

$$\begin{aligned} p(C_{\text{obs}}^{1:n}) &= \int_{-\infty}^{\infty} \mathcal{L}(K; C_{\text{obs}}^{1:n}) p(K) dK \\ &= \frac{1}{K_{\text{max}}} \int_0^{K_{\text{max}}} \prod_{i=1}^n \phi(C_{\text{obs}}(t_i); C(t_i), \sigma^2) dK, \end{aligned} \quad (\text{A.5})$$

which converges since Eq. (A.2) is bounded and continuous on the closed interval  $K \in [0, K_{\text{max}}]$ .

Finally, we arrive at the posterior through substitution of Eq. (A.3), Eq. (A.4), and Eq. (A.5) into Bayes' Theorem, Eq. (A.1),

$$p(K | C_{\text{obs}}^{1:n}) = \begin{cases} A \prod_{i=1}^n \phi(C_{\text{obs}}(t_i); C(t_i), \sigma^2), & \text{if } K \in [0, K_{\text{max}}], \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A.6})$$

where

$$\frac{1}{A} = \int_0^{K_{\text{max}}} \prod_{i=1}^n \phi(C_{\text{obs}}(t_i); C(t_i), \sigma^2) dK, \quad (\text{A.7})$$

which is Eq. (3) from the main text.

As stated in the main text, Eq. (A.6) assumes  $C(0)$  is fixed. To extend the model to capture uncertainty in  $C(0)$  we define  $C(0)$  to be a Gaussian random variable with mean  $\mu_0$  and variance  $\sigma_0^2$ . The PDF is  $p(C(0)) = \phi(C(0); \mu_0, \sigma_0^2)$ . In this case, the logistic growth curve, forming the means of the observations, depends on this random variable. Therefore, consider the joint distribution

$$\begin{aligned} p(K, C(0) | C_{\text{obs}}^{1:n}, \mu_0, \sigma_0^2) &= p(K | C(0), C_{\text{obs}}^{1:n}) p(C(0) | C_{\text{obs}}^{1:n}, \mu_0, \sigma_0^2) \\ &= p(K | C(0), C_{\text{obs}}^{1:n}) \phi(C(0); \mu_0, \sigma_0^2), \end{aligned} \quad (\text{A.8})$$

where  $p(K | C(0), C_{\text{obs}}^{1:n})$  is simply Eq. (A.6) with the dependence on  $C(0)$  made explicit. The desired posterior is a marginal density of Eq. (A.8), that is,

$$p(K | C_{\text{obs}}^{1:n}, \mu_0, \sigma_0^2) = \int_{-\infty}^{\infty} p(K, C(0) | C_{\text{obs}}^{1:n}, \mu_0, \sigma_0^2) dC(0). \quad (\text{A.9})$$

Substitution of Eq. (A.8) into Eq. (A.9) results in Eq. (4) in the main text. If expanded, the posterior is

$$p(K | C_{\text{obs}}^{1:n}, \mu_0, \sigma_0^2) = A \int_0^{K_{\text{max}}} \phi(C(0); \mu_0, \sigma_0^2) \prod_{i=1}^n \phi(C_{\text{obs}}(t_i); C(t_i), \sigma^2) dC(0),$$

for  $K \in [0, K_{\text{max}}]$  and  $p(K | C_{\text{obs}}^{1:n}, \mu_0, \sigma_0^2) = 0$  otherwise.

## Limiting uncertainty

We now consider the posteriors Eq. (A.6) and Eq. (A.9) in the limit as the observation times become infinitely large, that is  $t_i \rightarrow \infty$  for  $i \in [1, \dots, n]$ . First, note that since  $\lim_{t_i \rightarrow \infty} C(t_i) = K$ , it is also true that,

$$\lim_{t_i \rightarrow \infty} \phi(C_{\text{obs}}(t_i); C(t_i), \sigma^2) = \phi(C_{\text{obs}}^i; K, \sigma^2). \quad (\text{A.10})$$

That is, all observations are independent, identically distributed Gaussian random variables with mean  $K$  and variance  $\sigma^2$  in the limit. Since  $\phi(C_{\text{obs}}^i; K, \sigma^2) = \phi(K; C_{\text{obs}}^i, \sigma^2)$ , then we extend the domain of  $K$  to  $K \in (-\infty, \infty)$  and obtain

$$\lim_{(t_1, \dots, t_n) \rightarrow (\infty, \dots, \infty)} p(K | C_{\text{obs}}^{1:n}) = \frac{\prod_{i=1}^n \phi(K; C_{\text{obs}}^i, \sigma^2)}{\int_{-\infty}^{\infty} \prod_{i=1}^n \phi(K; C_{\text{obs}}^i, \sigma^2) dK}. \quad (\text{A.11})$$

Note that for any two Gaussian PDFs  $\phi(X; \mu_1, \sigma_1)$  and  $\phi(X; \mu_2, \sigma_2)$  it can be shown that

$$\phi(X; \mu_1, \sigma_1) \phi(X; \mu_2, \sigma_2) \propto \phi\left(X; \frac{\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right). \quad (\text{A.12})$$

Through some tedious algebraic manipulations, we apply Eq. (A.12) to Eq. (A.11) to obtain

$$\lim_{(t_1, \dots, t_n) \rightarrow (\infty, \dots, \infty)} p(K | C_{\text{obs}}^{1:n}) = \phi\left(K; \frac{1}{n} \sum_{i=1}^n C_{\text{obs}}^i, \frac{\sigma^2}{n}\right). \quad (\text{A.13})$$

Note that equality holds in Eq. (A.13) because the constants of proportionality in Eq. (A.12) cancel out through Eq. (A.11).

Similar steps are applied to Eq. (A.9) to obtain

$$\begin{aligned} \lim_{(t_1, \dots, t_n) \rightarrow (\infty, \dots, \infty)} p(K | C_{\text{obs}}^{1:n}, \mu_0, \sigma_0^2) &= \int_{-\infty}^{\infty} \phi\left(K; \frac{1}{n} \sum_{i=1}^n C_{\text{obs}}^i, \frac{\sigma^2}{n}\right) p(C(0)) dC(0) \\ &= \phi\left(K; \frac{1}{n} \sum_{i=1}^n C_{\text{obs}}^i, \frac{\sigma^2}{n}\right) \int_{-\infty}^{\infty} p(C(0)) dC(0) \\ &= \phi\left(K; \frac{1}{n} \sum_{i=1}^n C_{\text{obs}}^i, \frac{\sigma^2}{n}\right) \times 1 \\ &= \lim_{(t_1, \dots, t_n) \rightarrow (\infty, \dots, \infty)} p(K | C_{\text{obs}}^{1:n}). \end{aligned} \quad (\text{A.14})$$

This results in Eq. (5) of the main text.

It is important to note that in the derivation of Eq. (A.13) and Eq. (A.14) we have extended the domain of  $K$  to be the entire real number line. The prior in this case is actually improper, thus, for  $K \in (-\infty, \infty)$ , the posteriors Eq. (A.6) and Eq. (A.9) are not guaranteed to be PDFs. However, in the limiting case when  $t_i \rightarrow \infty$  for  $i \in [1, \dots, n]$ , the result is a PDF. Therefore, this extension is justified so long as  $K/\sigma \gg 1$ . If this is not satisfied, then this analysis could be completed using a truncated Gaussian prior.

## Short time uncertainty quantification

Here we derive Eq. (7) of the main text. The approximation uses the delta method to give an approximate variance for Eq. (A.6), that is the uncertainty in  $K$  for  $T < \infty$  and  $n = 1$ .

The posterior for  $n = 1$  is

$$p(K \mid C_{\text{obs}}(T)) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(C_{\text{obs}}(T) - C(T))^2 / (2\sigma^2)}. \quad (\text{A.15})$$

Thus, we can treat  $C(T)$  as a Gaussian random variable with mean  $C_{\text{obs}}(T)$  and variance  $\sigma^2$ . Given a realization of  $C(T)$  we can obtain from the logistic growth solution (Eq. (2) in main text)

$$K = f(C(T)) = \frac{C(T)C(0)(1 - e^{-\lambda T})}{C(0) - C(T)e^{-\lambda T}}. \quad (\text{A.16})$$

The delta method is an approximation technique for obtaining approximate moments of a distribution of a random variable that is a function of another [28]. The idea is to consider the Taylor expansion about the mean of  $C(T)$ ,

$$f(C(T)) = f(C_{\text{obs}}(T) + \delta) = \sum_{k=0}^{\infty} \frac{\delta^k}{k!} \frac{d^k f}{dC^k} [C_{\text{obs}}(T)].$$

From this expansion we obtain

$$\text{Var}[K] = \text{Var}[f(C(T))] = \text{Var}\left[\sum_{k=0}^{\infty} \frac{\delta^k}{k!} \frac{d^k f}{dC^k} [C_{\text{obs}}(T)]\right].$$

For small  $\delta$  we have

$$\text{Var}[K] = \text{Var}\left[f(C_{\text{obs}}(T)) + \delta \frac{df}{dC} [C_{\text{obs}}(T)] + \mathcal{O}(\delta^2)\right].$$

Recall that  $\text{Var}[a] = 0$ ,  $\text{Var}[X + a] = \text{Var}[X]$  and  $\text{Var}[aX] = a^2 \text{Var}[X]$  for any random variable,  $X$ , and constant  $a$ . Therefore,

$$\begin{aligned} \text{Var}\left[f(C_{\text{obs}}(T)) + \delta \frac{df}{dC} [C_{\text{obs}}(T)] + \mathcal{O}(\delta^2)\right] &= \text{Var}\left[\delta \frac{df}{dC} [C_{\text{obs}}(T)] + \mathcal{O}(\delta^2)\right] \\ &= \left(\frac{df}{dC} [C_{\text{obs}}(T)]\right)^2 \text{Var}[C(T)] + \mathcal{O}(\mu_4), \end{aligned} \quad (\text{A.17})$$

where  $\mu_4$  is the fourth central moment of the distribution of  $C(T)$ . The approximation is appropriate provided  $|\delta|$  is sufficiently small, that is  $|C(T) - C_{\text{obs}}(T)|$  is sufficiently small. This is achieved if  $C_{\text{obs}}(T)/\sigma \gg 1$ . The derivative of  $f$  with respect to  $C$  is

$$\frac{df}{dC} = \frac{C(0)^2(1 - e^{-\lambda T})}{[C(0) - C(T)e^{-\lambda T}]^2}. \quad (\text{A.18})$$

Therefore, by substitution of Eq. (A.18) into Eq. (A.17), we arrive at the approximation

$$\text{Var}[K] \approx \frac{C(0)^4(1 - e^{-\lambda T})^2}{[C(0) - C_{\text{obs}}(T)e^{-\lambda T}]^4} \sigma^2. \quad (\text{A.19})$$

There are two practical uses for Eq. (A.19). First, given an population density observation,  $C_{\text{obs}}(T)$ , taken at time  $T$ , one can calculate

$$\epsilon = \frac{C(0)^4(1 - e^{-\lambda T})^2}{[C(0) - C_{\text{obs}}(T)e^{-\lambda T}]^4}.$$

Here,  $\epsilon$  provides a relative uncertainty compared to the limiting best case  $\sigma$ . Therefore, the test provides a measure of how close to optimal the assay is. Furthermore, if the observation error is known then this sample approach provides the uncertainty in  $K$  after the observation. Second, given the upper limit of the prior  $K_{\max}$ , as in Eq. (A.4), one may consider the function

$$h(T) = \frac{C(0)^4(1 - e^{-\lambda T})^2}{[C(0) - K_{\max}e^{-\lambda T}]^4},$$

to identify how large  $T$  must be to be close enough to the limiting posterior, i.e.,  $\epsilon \approx 1$ . If this value of  $T$  is too large for practical purposes, then it indicates more observations should be taken. It is important to note that the approximation does require  $C_{\text{obs}}(T)/\sigma \gg 1$ . In practice, this may not hold for early time,  $T \ll 1/\lambda$ . In such a case, Eq. (A.19) tends to be an *underestimate*. Because of this, the result is still useful to decide if increasing  $n$  is valuable or not.

## Proliferation data

Table 1 presents the data used to inform Fig. 1, Fig. 2 and Fig. 3 in the main text. The data are derived from Jin et al. [19].

Table 1: Cell density data					
time (h)	0	6	12	18	24
cell density (cells/ $\mu\text{m}^2$ )	$3.1 \times 10^{-4}$	$3.8 \times 10^{-4}$	$5.2 \times 10^{-4}$	$5.9 \times 10^{-4}$	$7.8 \times 10^{-4}$

## References

- [1] Abercrombie, M. 1970. Contact inhibition in tissue culture. *In Vitro*, 6:128–142.
- [2] Levine, E. M., Y. Becker, C. W. Boone, and H. Eagle. 1964. Contact inhibition, macromolecular synthesis, and polyribosomes in cultured human diploid fibroblasts. *Proc. Natl. Acad. Sci. U. S. A.*, 53(2):350–356.
- [3] Liu, L, B. Sun, J. N. Pedersen, J. N. Pedersen, K-M. A. Yong, R. H. Getzenberg, H. A. Stone, and R. H Austin. 2011. Probing the invasiveness of prostate cancer cells in a 3D microfabricated landscape. *Proc. Natl. Acad. Sci. U. S. A.*, 108:6853–6856.
- [4] Abercrombie, M. 1979. Contact inhibition and malignancy. *Nature*, 281:259–262.
- [5] Puliafito, A., L. Hufnagel, P. Neveu, S. Streichan, A. Sigal, D. .K. Fygenson, and B. I. Shraiman. 2012. Collective and single cell behavior in epithelial contact inhibition. *Proc. Natl. Acad. Sci. U. S. A.*, 109(3):739–744.
- [6] Maini, P. K., D. L. S. McElwain, and D. I. Leavesley. 2004. Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. *Tissue Eng.*, 10(3–4):475–482.
- [7] Maini, P. K., D. L. S. McElwain, and D. I. Leavesley. 2004. Travelling waves in a wound healing assay. *Appl. Math. Lett.* 17(5):575–580.

- [8] Treloar, K. K., M. J. Simpson, D. L. S. McElwain, and R. E. Baker. 2014. Are *in vitro* estimates of cell diffusivity and cell proliferation rate sensitive to assay geometry? *J. Theor. Biol.* 356:71–84.
- [9] Tsoularis, A., and J. Wallace. 2002. Analysis of logistic growth models. *Math. Biosci.*, 179(1):21–55.
- [10] Enderling, H., and M. A. J. Chaplain. 2014. Mathematical modeling of tumor growth and treatment. *Curr. Pharm. Des.*, 20(30):4934–4940.
- [11] Gerlee, P. 2013. The model muddle: In search of tumor growth laws. *Cancer Res.*, 73(8):2407–2411.
- [12] Scott, J. G., D. Basanta, A. R. A. Anderson, and P. Gerlee. 2013. A mathematical model of tumor self-seeding reveals secondary metastatic deposits as drivers of primary tumour growth. *J. R. Soc. Interface*, 10(82):20130011.
- [13] Treloar, K. K., M. J. Simpson, P. Haridas, K. J. Manton, D. I. Leavesley, D. L. S. McElwain, and R. E. Baker. 2013. Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies. *BMC. Syst. Biol.*, 7:137.
- [14] Sherratt, J. A., and J. D. Murray. 1990. Models of epidermal wound healing. *Proc. R. Soc. Lond. [Biol]*, 241(1300):29–36.
- [15] Simpson, M. J., B. J. Binder, P. Haridas, B. K. Wood, K. K. Treloar, D. L. S. McElwain, and R. E. Baker. 2013. Experimental and modelling investigation of monolayer development with clustering. *Bull. Math. Biol.*, 75(5):871–889.
- [16] Chen, K., Q. Liu, L. L. Tsang, Q. Ye, H. C. Chan, Y. Sun, and X. Jiang. 2017. Human MSCs promotes colorectal cancer epithelial-mesenchymal transition and progression via CCL5/-catenin/Slug pathway. *Cell Death Dis.*, 8:e2819.
- [17] Delarue, M., F. Montel, D. Vignjevic, J. Prost, J.-F. Joanny, and G. Cappello. 2014. Compressive stress inhibits proliferation in tumor spheroids through a volume limitation. *Biophys. J.*, 107(8):1821–1828.
- [18] Beaumont, K. A., N. Mohana-Kumaran, and N. K. Haass. 2014. Modeling melanoma *in vitro* and *in vivo*. *Healthcare*, 2:27–46.
- [19] Jin, W., E. T. Shah, C. J. Penington, S. W. McCue, P. K. Maini, and M. J. Simpson. 2017. Logistic proliferation of cells in scratch assays is delayed. *Bull. Math. Biol.*, 79(5):1028–1050.
- [20] Sarapata, E. A., and L. G. de Pillis. 2014. A comparison and catalog of intrinsic tumor growth models. *Bull. Math. Biol.*, 76(8):2010–2024.
- [21] Huang, F.-T., W.-F. Chen, Z.-Q. Gu, Y.-Y. Zhuang, C.-Q. Li, L.-Y. Wang, J.-F. Peng, Z. Zhu, X. Luo, Y.-H. Li, H.-R. Yao, and S.-N. Zhang. 2017. The novel long intergenic noncoding RNA UCC promotes colorectal cancer progression by sponging miR-143. *Cell Death Dis.*, 8: e2778.

- [22] Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian Data Analysis*. CRC Press, Florida.
- [23] Liepe, L., S. Filippi, M. Komorowski, and M. P. H. Stumpf. 2013. Maximizing the information content of experiments in systems biology. *PLOS Comp. Biol.*, 9(1):e1002888.
- [24] Silk, D., P. D. W. Kirk, C. P. Barnes, T. Toni, and M. P. H. Stumpf. 2014. Model selection in systems biology depends on experimental design. *PLOS Comp. Biol.*, 10(6):e1003650.
- [25] Vanlier, J., C. A. Tiemann, P. A. J. Hilbers, and N. A. W. van Riel. 2012. A Bayesian approach to targeted experiment design. *Bioinformatics*, 28(8):1136–1142.
- [26] Vanlier, J., C. A. Tiemann, P. A. J. Hilbers, and N. A. W. van Riel. 2014. Optimal experiment design for model selection in biochemical networks. *BMC Syst. Biol.*, 8(1):20.
- [27] Browning, A. P., S. W. McCue, and M. J. Simpson. 2017. A Bayesian computational approach to explore the optimal duration of a cell proliferation assay. *Bull. Math. Biol.*, 79: 1888-1906.
- [28] Ang, A., and W. H. Tang. 2007. *Probability Concepts in Engineering*. Wiley, New York.