# Supporting Information

# Experimental evidence for bistable bacteria neutrophil interactions and potential clinical implications for neutropenic patients.

Roy Malka[1,2], Baruch Wolach[3,4], Ronit Gavrieli[3]

Eliezer Shochat[5], Vered Rom-Kedar[1,6]

[1]Department of Computer Science & Applied Mathematics,

Weizmann Institute of Science, Rehovot, Israel.

[2] Current address: Center for Systems Biology, Massachusetts General Hospital, Boston, MA 02114 and Department of Systems Biology, Harvard Medical School, Boston, MA, 02115.

[3] The Laboratory for Leukocyte Function , Meir General Hospital, Kfar-Sava, Israel

[4] The Department of Pediatrics, Meir General Hospital, Kfar-Sava, Israel.

[5]Pharma Development, Hoffmann-La Roche, Basel, Switzerland.

[6]The Estrin family chair of computer science and applied mathematics.

We first supply more details with regards to the goodness of fit of the mathematical model (1) of Methods to the experimental data. We then complete the information needed for simulating eq. (2) of Methods to produce Fig 4 (presenting the functional form of $I_{N,G}(t)$ and the complete set of parameter values used in the first two equations). We end with providing full details regarding the sign test.

## Model

We first compare in detail the goodness of fit and the prediction error distributions of the three variants to model (1) of Methods (the relevant formula are presented here again for the readers convenience):

$$\frac{dB(t)}{dt} = \rho B(t) - \frac{\alpha N B(t)}{1 + \gamma B(t) + \eta N}. \tag{1}$$

## Estimation Methodology

The proposed models are all deterministic: given a set of parameters $\mu$ and the initial data of an experiment $X = (B(0), N)$, a deterministic model predicts that the bacterial concentration after 60 minutes is $\hat{B}(X, \mu)$ whereas the experimental results produce some given concentration $B(X)$. It is assumed that the source of errors in the model's prediction is associated with the measurement process. Due to the nature of the modeled process (bacterial multiplication) and the nature of the measurements (based on dilutions)

the measurement errors are expected to be multiplicative in nature:

$$B(X) = \hat{B}(X, \mu) \times (1 + z),$$

where $z$, the prediction error or residual, is distributed according to some unknown distribution. Indeed, we observe that the variability in the measurements is proportional to values of the measurements: typical duplicates yield measurements such as $b_1 = 10^6$ and $b_2 = 2 \times 10^6$ cfu, and, for low concentrations $b_1 = 10^4$ and $b_2 = 2 \times 10^4$ cfu (an increased variance for an increased initial concentration of bacteria).

Given $(B(X), X, \mu)$, the prediction error may be extracted:

$$z = 1 - \frac{B(X)}{\hat{B}(X, \mu)}.$$

Hereafter, we use the following notation: $z$ is reserved for the general discussion. When the prediction $\hat{B}(X, \mu)$ is formed by the linear model (setting $\gamma = \eta = 0$ in (1) and fitting $\alpha$) we denote the prediction error by $z^L$, whereas $z^{NL2}$ and $z^{NL}$ denote, respectively, the prediction errors for the nonlinear model with no neutrophil interference (setting $\eta = 0$ in (1) and fitting $\alpha, \gamma$) and the fully non-linear model (1).

To examine the process reliability, three appropriate goodness-of-fit (GF) measures are defined:

$$LS = \frac{1}{n}\sum_{i=1}^{n} z_i^2 \quad \text{- Least-Squares}$$

$$LS_a = \frac{1}{n}\sum_{i=1}^{n} |z_i| \quad \text{- Double Exponential}$$

$$LS_b = \frac{1}{n}\sum_{i=1}^{n} \ln(1+\frac{1}{2}z_i^2) \quad \text{- Lorentzian}$$

Note that we do not assume that the data is indeed normally, double exponential or Lorentzian distributed (in such a case the estimator will be the maximum likelihood estimator). We use these as a goodness-of-fit measure to generate a robust estimation [1].

The parameter fitting amounts now to a minimization problem of the GF measure that is solved numerically using the Matlab function 'fminsearch', which implements the multidimensional downhill simplex method (see e.g., [1, 2]). In all cases we find that the three GF measures lead, to the first significant digit, to the same minimizing parameter values. For the data adopted from [3] we utilize their estimate for the bacterial growth at the exponential phase and then use the same procedure as above to fit the kill term parameters.

## Growth Curve Parameter Estimation

The parameters associated with the bacteria natural dynamics are fitted to the data of growth calibration experiments for different time points and to

the control experiments at $N = 0$ for the time point $t = 60$ min. Interestingly, the three-parameter fit to the growth curve (fitting $(r, d, \beta)$) did not produce better results than the one parameter linear fit, indicating that the bacteria are still in the exponential phase in these experiments; thus we set $\beta = 0$.

## Kill-Term Parameter Estimation

The bacterial dynamics are assumed to be fixed in all experiments; thus, here we fit only the parameters relevant to the kill term. All three measures lead to the same conclusion - the best fit is found when the saturated non-linear model is considered, and the parameters that provide the best fit (using the $LS_b$ measure) to the data are listed in Table 2 of the manuscript. All the measures reveal consistency in the improvements from one to two and to three parameters.

Lilliefors' and Jarque-Bera tests for normally distributed data cast doubt on the normality of $z$ (in particular, the normality hypothesis is rejected with significance 0.05). Thus, in Table 2 we present only the parameters fitted using the $LS_b$ measure. This choice has no qualitative impact on the conclusions.

## Goodness-of-fit via hypothesis testing

The goodness-of-fit measures suggest that the bistable models are better than the linear (neutrophil-threshold) model. To examine the significance of these

improvements we use nonparametric statistical tests. We employ the the Kolmogorov-Smirnov test (KS) [4] and the Mann-Whitney-Wilcoxon test [5] to compare the distribution of the prediction errors obtained using the linear and the non-linear models. Figure 1(a) presents the estimated[1] probability distribution of the prediction error for the three models ($z^L$-linear in red (fitted $\alpha$), $z^{NL2}$ - non-linear with no interference (fitted $\alpha, \gamma$) in green, and $z^{NL}$ fully non-linear (fitted $\alpha, \gamma, \eta$) in blue).

Notably, the mean of the absolute value of the predicted error in the linear model $|z^L|$ is three times larger than that of the non-linear one $|z^{NL}|$. Fig 1(c) shows the empirical distributions of $w_1 = |z^L| - |z^{NL}|$, $w_2 = |z^L| - |z^{NL2}|$ and $w_3 = |z^{NL2}| - |z^{NL}|$. We see that the signed error difference $w_1$ has a mean of about 0.35 (positive) . Thus, we conclude that the prediction errors of the linear model are indeed bigger than those of the nonlinear models.

**Hypothesis testing:** To further substantiate the claim that the fully nonlinear model is superior, we propose the following three one-sided tests:

$H_0$**:** $|z^L|$ and $|z^{NL}|$ are drawn from the same distribution ($F^L(|z|) = F^{NL}(|z|)$).

$H_1$**:** alternatively $F^L(|z|) < F^{NL}(|z|)$.

$G_0$**:** $|z^L|$ and $|z^{NL2}|$ are drawn from the same distribution ($F^L(|z|) = F^{NL2}(|z|)$).

$G_1$**:** alternatively $F^L(|z|) < F^{NL2}(|z|)$.

[1]The pdf estimate is obtained using Gaussian kernel estimation, as implemented in Matlab, and its sole use is visualization.

$I_0$**:** $|z^{NL2}|$ and $|z^{NL}|$ are drawn from the same distribution ($F^{NL2}(|z|) = F^{NL}(|z|)$).

$I_1$**:** alternatively $F^{NL2}(|z|) < F^{NL}(|z|)$.

where $F^L$, $F^{NL2}$ and $F^{NL}$ are the cumulative distribution functions corresponding to the relevant models. The null hypothesis $H_0$ is rejected with significance $= 0.001$ using the nonparametric one-sided Kolmogorov-Smirnov (KS) test [1, 4]. The null hypotheses $G_0$ and $I_0$ are rejected with significance $= 0.05$. In all the cases the alternative hypothesis is accepted. The same conclusion holds when applying the Mann-Whitney-Wilcoxon test [5] with the same significance. The implication of the alternative hypothesis is that the probability of $\{|z| < z_0\}$ is significantly larger under the assumption, e.g., prediction obtained using the non-linear model versus the linear one (H-test). Put differently, there is more probability-mass on the small prediction errors in the non-linear case than in the linear case (see Figures 1(b)-1(c)).

The above statements generate an order among the non-linear-3 model (fitting $(\alpha, \gamma, \eta)$), the non-linear-2 model (fitting $(\alpha, \gamma)$) and the linear model (fitting $\alpha$). Using $\eta > 0$ gives much better significance ($p < 0.001$) than using $\eta = 0$ ($p < 0.05$). Moreover, the saturated model is significantly better than the model with $\eta = 0$, with significance level 0.05. This establishes that the kill term should have saturation in both the bacterial and the neutrophil concentrations.

*Technical remark 1:* The statistical tests require the independence of the

data sets, $z^L$ and $z^{NL}$, whereas here these data sets are produced from the same experimental data. To overcome this problem we use the standard trick of randomly dividing the data into two parts and use different data parts to produce the prediction error of each model. This process of random partition was repeated 10,000 times. In the $H$ case, the non-linear model (with $\eta > 0$) was always better, with significance level of 0.001 (actual p-values are $10^{-6}$ or less). Also this model was always better than the second non-linear model (with $\eta = 0$), i.e., the $I$ test, with a significance level of 0.05 (actual values rarely exceed 0.01 – it holds with significance level of 0.01 for 98% of the simulations). The above holds whether part one is used to generate prediction errors for the one model (e.g., linear) or the other model (e.g., non-linear). The $G$ test, namely the second non-linear model (with $\eta = 0$), is better than the linear model with significance level of 0.05, though only in one of the two parts, while it holds for both parts of the data, as in the H and I tests, in 87% of the simulations (the null hypothesis was never accepted in both parts).

*Technical remark 2:* The Mann-Whitney-Wilcoxon test provides similar conclusions. In particular, in the worst case of the $G$ test with significance level of 0.05, it holds for the two parts in 97% of the simulations.

# Parameters of the mathematical model for the severe chemotherapy induced neutropenia scenario.

Fig 4 of the manuscript is produced by simulating Eq. (2) of Methods:

$$\begin{cases} \frac{dN}{dt} = \overbrace{I_N(t)\frac{k_G \ + \ k_{NEF} \ G}{k_G \ + \ G}}^{\text{influx of N from the bone marrow}} \ - \ \overbrace{D_N \cdot N}^{\text{N clearance}} \ . \\ \\ \frac{dG}{dt} = \overbrace{I_G(t)}^{\text{Influx of G by injections}} + \overbrace{\frac{B_G^{max}}{1 \ + \ N/k_N} - D_G^r \cdot G - \frac{D_G^n \cdot N}{k_N \ + \ N} \cdot G}^{\text{physiological G-CSF dynamics}} \\ \\ \frac{dB}{dt} = \rho B + s - \frac{\alpha N B}{1+\gamma b B+\eta N} \end{cases} \tag{2}$$

Here we complete the details needed to perform these simulations. The functional form of $I_{N,G}(t)$ that appear in Eq. (2) are:

$$\begin{aligned} I_N(t) &= B_{nadir} + (B_N - B_{nadir})\times \\ &\quad \left(1 - \frac{\tanh(\beta_1(t - T_{start})) - \tanh(\beta_2(t - T_{stop}))}{2}\right) \\ I_G(t) &= \sum_{j=1}^{10} \ \lambda \ \frac{dose_G}{v_d} \ e^{-\lambda \max\{0,(t-T_{gstart}-jT_f)\}}, \ t \geq T_{gstart} \end{aligned}$$

The first form reflects the rapid destruction and recovery of the bone-marrow ability to supply neutrophils to the blood due to a chemotherapy treatment. The second form corresponds to the common pharmacokinetics associated with standard 10 daily G-CSF injections. These functional forms fit well published clinical data sets [6]. The parameters used in these terms and in the other terms that appear in the first two equations that govern the neutrophils and G-CSF dynamics are presented in Table 1. The parameters for the third equation, governing the bacterial dynamics, are exerted from the fitted parameters to the in-vitro experiments as presented in Table 2 of the manuscript. The constant bacterial influx is as indicated. See [7] for more details regarding this combined model.

## Details regarding the Sign test

Here we provide more details regarding the implementation of the sign test to the bacteria-neutrophil in-vitro experimental data.

Due to the finite number of points in our experiments, the set of possible slanted straight lines that provide different classifications is finite. We test the null hypothesis on a large portion of the relevant lines in the $X = (N, B(0))$ plane by taking a grid over the lines' slopes and their N-intercepts. Under the fair coin hypothesis, the binomial distribution describes the success of the classification and provides the power of the test. We find that the null hypothesis is rejected for all the lines with slopes in the range $(\tan 55.8^o, \tan 88.4^o)$ and with an N-intercept point located in the range

$(0, 7.5 \times 10^5)$ with significance of less than 0.01, where the highest $p$-value is less than $p_{max} = 0.001016$. In fact, the best lines have $p$-value of the order of $10^{-7}$. These best lines have successfully classified more than 85% of the data points (the lines that have larger $p-$values that are close to $p_{max}$ have success rate of at least 75%). The above statistics are found for the combined data set of the 4 subjects. When each subject is considered separately, the results are usually even better: the success in classification of the best lines are 93%, 95%, 95%, and 91% with p-values $8.54\times10^{-4}, 1.9\times10^{-5}, 5.24\times10^{-6}$, and $6.03 \times 10^{-5}$ for patients P1-P4 respectively. The nonlinear BNCs that emerge from the fitted mathematical model have about 80-90% success in the classification, more details appear in the parameter fitting section.

On the other hand, when we repeat the above analysis for nearly vertical lines, in particular for lines with slopes in the range $(\tan 89^o, \infty) \cup (-\infty, \tan 120^o)$, we find that the null hypothesis cannot be rejected with any significance. This result means that the same distribution of increasing/decreasing data points with respect to vertical lines could be possibly achieved by a fair coin toss. The same conclusion (with smaller p-values) applies to lines with strictly negative slopes.

Thus, the statistical tests reject the neutrophil-threshold model that corresponds to a vertical classifier and the monostable behavior that corresponds to a classifier with a negative slope. The statistical test does support the existence of a slanted critical line with a positive slope as in the bistable and the ratio-dependent behaviors. Even though we did not consider here nonlinear

classifiers, it is clear from the data that any classifier that corresponds to a monotonically decreasing BNC is a worse classifier than a vertical one, and vertical lines were rejected.

Finally, Figure 2 B presents the sign diagram of the *S. epidermidis* data extracted from [3], where neutrophils and *S. epidermidis* were added to a fibrin gel which simulates human tissue. The bacterial concentration was then recovered from the gel after 90 minutes. We see several neutrophil concentrations at which low bacterial concentrations have the O sign and high bacterial concentrations have the + sign, hence bistability appears in a different experimental setting and with a different bacterial strain. However, here, the number of data points near the unstable branch is insufficient for obtaining definite conclusions from the sign-test analysis.

*Technical Remark:* We also test whether the position of the initial data points and the division between increasing and decreasing points in our data may skew the statistical analysis. To this aim we took, 2000 times, a random permutation over the symbols '+1/-1' at the experimental initial collective data points $N$ and $B(0)$. For less than 1% of the permutations we found a line for which the null hypothesis could be rejected with significance of 0.05. This numerical experiment supports the claim that it is highly improbable that the observed behavior can appear at random.

# References


[1] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.

[2] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9:112–147, 1998.

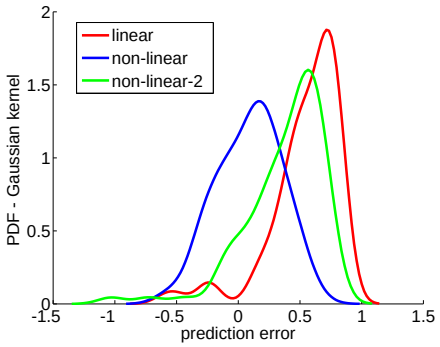
[3] Y. Li, A. Karlin, D.J. Loike, and C.S. Silverstein. Determination of the critical concentration of neutrophils required to block bacterial growth in tissues. *The Journal of Experimental Medicine*, 200(5):613–622, 2004.

[4] R. von Mises. *Mathematical theory of probability and statistics*. Academic Press, 1964.

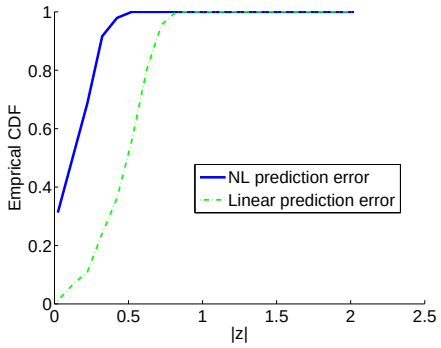
[5] R. V. Hogg and A. T. Craig. *Introduction to Mathematical Statistics*. Prentice-Hall, 5th edition, 1995.

[6] E. Shochat and V. Rom-Kedar. Novel strategies for g-csf treatment of high-risk severe neutropenia suggested by mathematical modeling. *Clinical Cancer Research*, 14(7):6354–6363, October 15 2008.

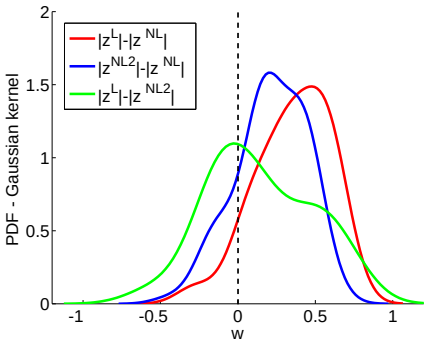
[7] R. Malka and V. Rom-Kedar. Bacteria–phagocytes dynamics, axiomatic modelling and mass-action kinetics. *Mathematical Biosciences and Engineering*, 8(2):475-502, 2011.




(a) Prediction error




(b) CDF: H hypothesis




(c) Directional prediction error

Figure 1: **Visualising the statistics. (a)** The estimated probability distribution of the prediction errors (approximated using Gaussian kernels) provides a visualization of the error distributions when the three different deterministic models are used to fit the data. Red: linear model, Blue: the full non-linear model (NL), Green: the non-linear model with bacterial saturation ($\gamma > 0$) and without neutrophil saturation ($\eta = 0$). Only the full non-linear model provides an error distribution that is nearly symmetric around zero. **(b)** The empirical CDF of $|z|$ for the linear (green) and non linear (blue) models provides a visualization of the H test. **(c)** The distribution of the difference (directional) between the linear prediction error and the non-linear prediction error demonstrates that the error in fitting the non-linear model is usually smaller than the error achieved by the linear fit.

| Parameter | Meaning | $\mu$ G-CSF | Units |
|---|---|---|---|
| $G(0)$ | $G$ initial value | 50 | $pg/ml$ |
| $N(0)$ | $N$ initial value | $3 \cdot 10^6$ | $\#cells/$ml |
| $B_G^{max}$ | $G$ max production rate | 4.861 | $pg/ml/minute$ |
| $k_N$ | $N$ dissociation constant | $0.5 \cdot 10^6$ | $cells/ml$ |
| $k_G$ | $G$ MM constant | $5 \cdot 10^3$ | $pg/ml$ |
| $D_G^n$ | $G$ elimination (by $N$) | 0.0048 | $1/minute$ |
| $D_G^r$ | $G$ elimination (renal) | 0.0042 | $1/day$ |
| $B_N$ | Basic normal flux of $N$ | $15 \cdot 10^6$ | $\#cells/ml/day$ |
| $k_{NEF}$ | Enhancement of $N$ flux by $G$ | 10 | $-$ |
| $D_N$ | $N$ clearance rate | 0.0021 | $1/minute$ |
| $T_{start}$ | Start of CT effect | 2880 (day 2) | $minute$ |
| $T_{stop}$ | Stop of CT effect | 23040 (day 16) | $minute$ |
| $\beta_1$ | Marrow depletion rate | $5.2 \times 10^{-4}$ | $1/minute$ |
| $\beta_2$ | Marrow recovery rate | $5.2 \times 10^{-4}$ | $1/minute$ |
| $AMC$ | Acute Marrow capacity | 0.15 | $-$ |
| $Bnadir$ | Marrow $N$ flux at nadir$=\frac{AMC \cdot N^* D_N}{k_{NEF}}$ | 156.25 | $\#cells/ml/minute$ |
| $\lambda$ | $G$ absorption rate | 0.0028 | $1/minute$ |
| $v_d$ | volume of $G$ distribution | 2300 | $ml$ |
| $w$ | Weight | 70 | $Kg$ |
| $dose_G$ | G-CSF dose in a single shot | $5 \times 10^6$ | $pg/Kg$ |
| $T_{gstart}$ | G-CSF injections start time | $T_{start} + 2880$ | $minute$ |
| $T_f$ | G-CSF injections period | 1440(daily) | $minute$ |

Table 1: The chosen Neutrophils-G-CSF model parameters are as in table 1 of [7], with the exception of the $AMC$ value that is slightly smaller (there AMC=0.2) and here only daily injections are considered. The duration of the chemotherapy, the AMC value and the starting day of the G-CSF therapy are slightly altered from the [6] values, all the other parameters are as in [6] (notice that we divide the daily rates of table 2 of the supplement of [6] by 1440 to obtain rates per minute). $N^* = 5 \cdot 10^6 cells/ml$ is a fixed scaling parameter.