Monte Carlo simulation-based approach to model the size distribution of metastatic tumors

Esha Maiti*
California High School, San Ramon, California 94583, USA
(Received 23 October 2011; revised manuscript received 16 December 2011; published 27 January 2012)

The size distribution of metastatic tumors and its time evolution are traditionally described by integrodifferential equations and stochastic models. Here we develop a simple Monte Carlo approach in which each event of metastasis is treated as a chance event through random-number generation. We demonstrate the accuracy of this approach on a specific growth and metastasis model by showing that it quantitatively reproduces the size distribution and the total number of tumors as a function of time. The approach also yields statistical distribution of patient-to-patient variations, and has the flexibility to incorporate many real-life complexities.

DOI: 10.1103/PhysRevE.85.012901 PACS number(s): 87.10.Rt, 87.19.xj

I. INTRODUCTION

Metastasis is the spreading of tumors from a primary tumor source that leads to the formation of secondary tumor colonies in different locations within the patient's body. Multiple metastases are a major problem that presents a severe challenge to the treatment of cancer [1,2]. Modeling the evolution of the size and numbers of metastatic tumors as a function of time can be useful toward detecting and treating early forms of cancer. The rate of metastasis depends on both the tumor size as well as the distribution of blood vessels (vasculature) around and within the metastasizing tumor [3,4]. Although advances in clinical imaging technology makes it possible to accurately measure the size of tumors or to get quantitative information on the vasculature [5], it is still difficult to observe very small tumors with current clinical techniques [6]. Thus, a mathematical model of tumor growth and metastasis can be invaluable.

Of the multitude of models published in the literature over the last decade [7–12], one of the earlier ones developed by Iwata, Kawasake, and Shigesada (IKS) [12] is simple and elegant. Their model makes two simple assumptions: (1) Any tumor (primary, secondary, tertiary, etc.) starts from a single cell and grows according to the Gompertz curve [13,14]:

$$x(t) = b^{1-e^{-at}}, \tag{1}$$

where $x(t)$ is the size of the tumor (expressed as the number of cells comprising the tumor) at time t, b is the maximum possible size of the tumor, and a is a rate constant; and (2) for any given tumor size x, the rate of metastasis, $\beta(x)$, is proportional to the degree of angiogenesis:

$$\beta(x) = mx^{\alpha}, \tag{2}$$

where m is a rate constant (colonization coefficient), and α is the fractal dimension of blood vessels connected to and infiltrating the tumor [15–17]. IKS then defined a size distribution of the metastatic colony $\rho(x,t)$ [where $\rho(x,t)dx =$ number of metastatic tumors of sizes between x and $x + dx$ at time t] and used the von Foerster equation [18] to describe its time evolution under the conditions imposed by Eqs. (1) and (2). An in-depth mathematical analysis of the IKS model has recently been performed by Barbolosi *et al.* [19].

Although the IKS approach is mathematically elegant and could even be solved analytically, it cannot be easily extended to more realistic situations, e.g., the growth law for the metastasized tumors could be organ-environment dependent and significantly different from the law governing the primary tumor growth. Even a single large tumor could have intrinsic heterogeneities, e.g., niches or compartments [20–23] that require different metastasis rates from different parts of the tumor. Additional complications could involve a time lag between the formation and shedding of a metastatic tumor from its primary source, a finite probability of its survival through the body's immune system, and so on [24]. To incorporate such complexities one would need significant modifications to the IKS model, which could be nontrivial. In this paper, we adopt a much simpler approach in which we treat metastasis as a probabilistic event and perform Monte Carlo simulations to determine the size and number evolution of all tumors as a function of time. Such approach possesses the flexibility to incorporate some of the real-life complexities mentioned above. For instance, if an extensive database of medical records of cancer patients could be analyzed to determine the probabilities of metastasis as a function of the organ of origin and the organ of spread, such knowledge could be incorporated in the Monte Carlo procedure described below. It would also be relatively straightforward to incorporate organ-specific growth rates, possible time lags of shedding, or finite probabilities of immune destruction of tumors. However, prior to including any of such complexities in the programming logic one first needs to validate the accuracy of the Monte Carlo approach by reproducing some of the results of the IKS model. Carrying out such validation is the main purpose of this paper. In the following section we describe the simulation procedure in more detail.

II. SIMULATION PROCEDURE AND VALIDATION

For a specific patient, we assume that a primary tumor appears as a single cell on day 0 and grows according to the Gompertz function, i.e., Eq. (1). Then we interpret Eq. (2) as a probability rate of metastasis of this tumor. Equation (2) describes a rate process that is continuous in time. For the purpose of numerical simulation we divide this continuous time into time intervals Δt such that the probability of metastasis within any given time interval is smaller than a small number ε. In the beginning, when the primary tumor is

*emaiti96@gmail.com

very small, the probability of its metastasis per day is low. In this regime we choose $\Delta t = 1$ day. However, when the size of the primary tumor becomes larger (e.g., 10^9 or so), we choose Δt such that $mx^\alpha \Delta t \leqslant \varepsilon$. Thus, a general time step is defined through the function

$$\Delta t = \min\left(1, \frac{\varepsilon}{mx^\alpha}\right). \tag{3}$$

From several test runs (see procedural details below) with different values of ε we determined that $\varepsilon = 0.05$ is a good choice for simulations of total time 5 years or less—smaller values of ε take longer to run (for the same total simulation time) and yet yield essentially identical results.

The simulation proceeds by increasing time through a time step Δt [given by Eq. (3)], computing the probability of metastasis $mx^\alpha \Delta t$, and drawing a random number. If the random number is between 0 and $mx^\alpha \Delta t$, a metastatic tumor of size one cell is created. This metastatic tumor then starts to grow according to the Gompertz function along with the primary tumor, which continues to produce more metastatic tumors according to the probability rate discussed above. Following a given period of observation (a few years), we count the number and sizes of all tumors: primary, secondary, tertiary, and so on. We repeat the above procedure for many other patients and average over all the patients to obtain a statistically averaged result.

In order to check that such numerical experiment is able to produce the same results as IKS, we first examined the cumulative size distribution and the total number of tumors as a function of time using the same parameters as IKS. The parameters used in our simulations were $a = 0.00286$ day^{-1}; $b = 7.3\times10^{10}$ cells; $m = 5.3\times10^{-8}$ day^{-1}; and $\alpha = 0.665$. Most of the simulations reported below were averaged over 200 patients (we checked that a larger number of patients essentially produces the same results).

Figure 1 displays the cumulative size distribution of tumors of size greater than 10^7 cells for four different times of observation, i.e., 1100, 1227, 1300, 1400 days. Given an

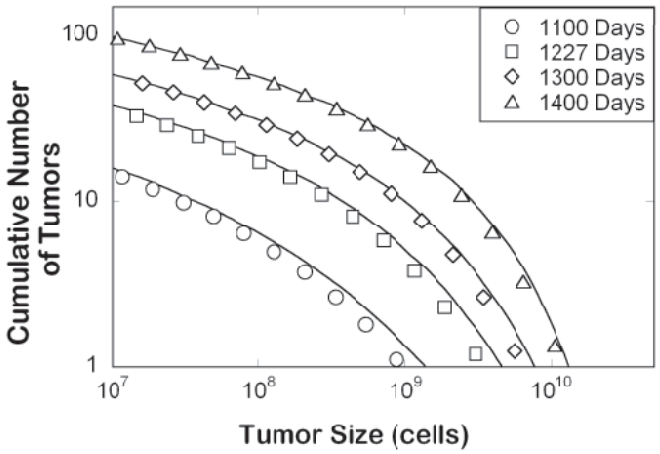

FIG. 1. Cumulative size distribution of tumors from our simulations (symbols) averaged over 200 patients for four different times of observation. The solid lines are results from Fig. 4 (upper) of IKS (Ref. [12]) corresponding to 432 days, 559 days, 632 days, and 732 days, respectively. (See text.)

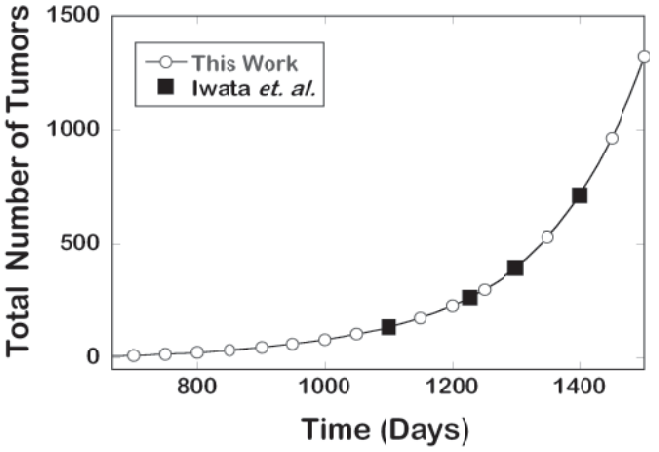

FIG. 2. Total numbers of tumors from our simulation averaged over 200 patients as a function of time, compared with the results from Iwata *et al.* (Ref. [12]).

estimated origination time of −668 days for the primary tumor in the IKS paper, these times correspond to 432, 559, 632, and 732 days in their work. Excellent agreement between our simulated distributions and the IKS results proves the accuracy and validity of our numerical method.

Figure 2 displays the total numbers of metastasized tumors (secondary and tertiary) as a function of time for the first 5 yr, along with IKS values for four different times during this period. As can be seen, the IKS values fall right on our curve, again showing the accuracy of our simulations.

III. ADDITIONAL RESULTS AND DISCUSSION

Figure 3 displays separately the secondary and tertiary tumor growth as a function of time. It also shows the corresponding numbers of larger ($>10^9$ cells, which are 1 cm^3 or larger) tumors in the body. From the results, we find that on an average the secondary tumors begin to develop at around ∼500 days, while the first tertiary tumors begin to develop at around ∼900 days. The bigger secondary and tertiary tumors begin to appear at around 1100 days and 1500 days, respectively. We notice from the graph that although the tertiary tumors appear later than the secondary tumors, they multiply faster and outnumber the secondary tumors within ∼3.6 yr [see Fig. 3(b)].

Noticing linear behavior of the number of secondary and tertiary tumors over certain time segments in the log-log plot [the two leftmost curves in Fig. 3(b)] we attempted to extract a power law behavior [25] of the total number of metastatic tumors as a function of time. From Fig. 3(b) it is evident that one cannot fit a single power law over the entire range of 0–5 yr. Instead, the early and later stages follow two different power laws: For the first 2 yr the metastatic tumors are entirely secondary in nature and follow the behavior $0.14t^{6.69}$ (t = time in years), while in the time frame of 3–5 yr it follows the behavior $0.01t^{8.36}$, the steeper exponent being due to the rapid proliferation of the tertiary tumors in this period [see Fig. 3(b)].

So far all of the above discussion focused on quantities *averaged* over a number of patients, i.e., 200. However, in

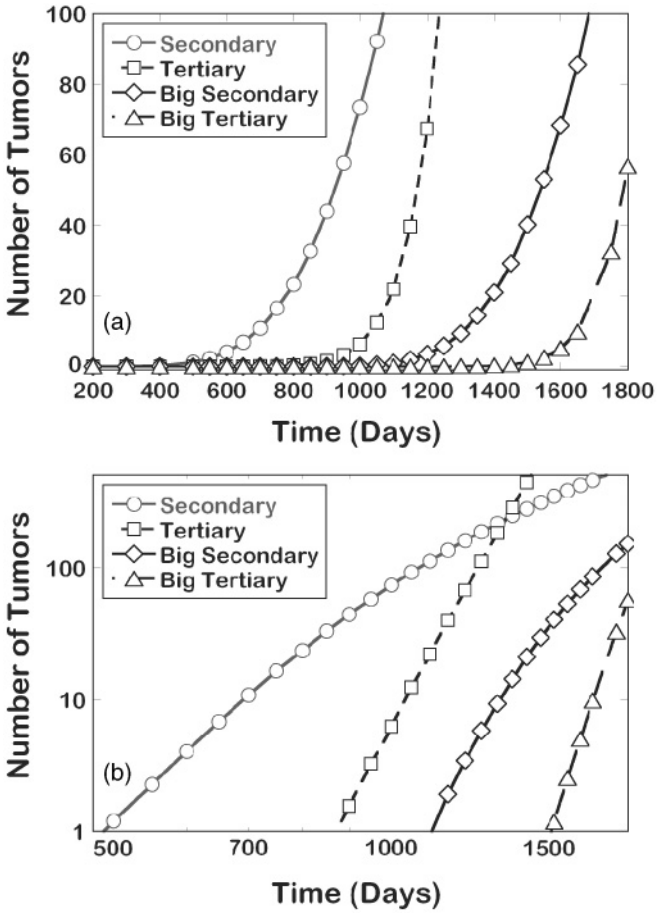

FIG. 3. A breakdown of the number of secondary and tertiary metastasized tumors as a function of time; (a) linear-linear plot; (b) log-log plot. The results displayed are from our numerical simulations averaged over 200 patients. Also shown are the number of corresponding large tumors of sizes $>10^9$ cells (1 cm^3 or larger).

order to obtain the statistical distribution of a given variable it is necessary to store the data for all individual patients and create a frequency distribution. As an illustration, we plot in Fig. 4 the distribution of the time at which the first metastasis for the primary and the secondary tumors occur. These results were obtained from 1200-day-long simulations on 5000 patients. Compared to averaged quantities (as in Figs. 1–3) where a much smaller number of patients (i.e., 200) was sufficient, a smooth statistical distribution mandated simulations over a much larger number (i.e., 5000). Both distributions are nearly normal with small negative skews, and could be accurately fit with the skewed normal distribution (SND) as defined by the

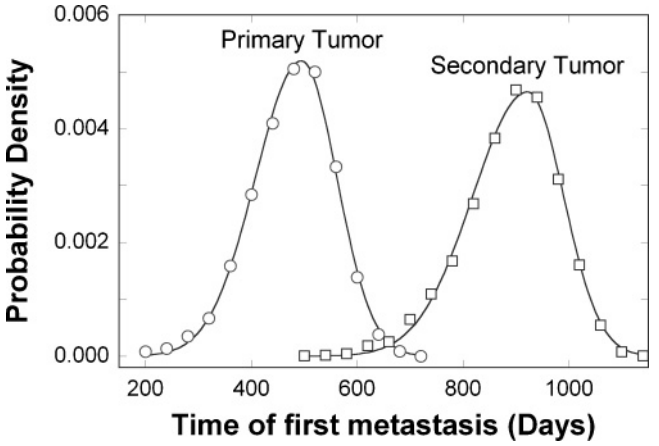

FIG. 4. Statistical distribution of the time at which the first metastasis of the primary and the secondary tumors occur. The results are from our simulations on 5000 patients. Solid curves are best fits using the skewed normal distribution (SND). Both curves are negatively skewed with skewness factors −0.13 and −0.21 for the primary and secondary curves, respectively. (See text and Table I).

probability distribution function [26]:

$$f(x) = \frac{(1-s)^2}{\sigma\sqrt{2\pi}} e^{-[(x-x_M)-s|x-x_M|]^2/2\sigma^2}, \qquad (4)$$

where s is a skew factor, x_M is the position of the peak (i.e., mode), and σ a measure of the distribution width. [Note that $s = 0$ corresponds to the regular normal distribution with mean x_M and standard deviation σ.] The SND fitting parameters for the two curves in Fig. 4 and the associated mean, standard deviation, and skewness are listed in Table I. From the values listed in this table it becomes clear that the secondary distribution has a bigger width and higher (i.e., more negative) skewness than the primary distribution. The negative skewness of both distributions can be attributed to higher uncertainty in metastasis rates when the tumor is small (as compared to a matured tumor when it is more certain to metastasize). Also the degree of uncertainty for the metastasis of a secondary tumor is higher because its origination itself involves uncertainties of metastasis of the primary tumor. This explains the wider peak and higher degree of (negative) skewness of the secondary peak as compared to the primary one.

With preventive care in mind, it is interesting to look at the earliest onset of metastasis. For a particular patient, primary metastasis occurred as early as 200 days within the formation of the primary tumor. In fact, as much as 5% of the patients had primary metastasis occur within 347 days. Similarly, in the most aggressive cases, secondary metastasis occurred within

TABLE I. Fit parameters and derived quantities for the statistical distributions of Fig. 4. The data were obtained by simulating 5000 patients and fitted using the skewed normal distribution (SND).

	SND fit parameters			Derived quantities from SND fit		
Tumor type	x_M	σ	s	Mean	Standard deviation	Skewness
Primary	494.5	75.5	−0.13	478.6	76.8	−0.20
Secondary	922.0	82.0	−0.21	893.1	86.4	−0.34

500 days of the formation of the primary tumor, with 5% of the patients seeing it occur within 742 days.

IV. SUMMARY

In summary, we show that simple Monte Carlo simulations can be very useful in predicting the evolution of size distribution of metastatic tumors. The accuracy of our approach is demonstrated through simulations on the IKS model [12] in which the same growth and metastasis rates are used for all tumors: primary, secondary, or tertiary. From patient-to-patient variations one can also obtain the statistical distribution of useful quantities, e.g., that of the time at which the first metastasis of the primary and secondary tumors occur, which display negatively skewed normal behavior. Future work will involve more complex (and realistic) situations in which the vasculature and growth rate laws of the metastatic tumors can vary depending upon their respective organ environments, and possible heterogeneities within single large tumors are accounted for.

[1] C. W. Evans, *The Metastatic Cell: Behavior and Biochemistry* (Chapman & Hall, London, 1991).
[2] M. M. Mareel, P. De Baetsellier, and F. M. Van Roy, *Mechanisms of Invasion and Metastasis* (CRC Press, Boca Raton, FL, 1991).
[3] G. Poste and I. J. Fidler, Nature **283**, 139 (1980).
[4] V. Schirrmacher, Adv. Cancer Res. **43**, 1 (1985).
[5] H. Ohishi *et al.*, J. Ultrasound Med. **17**, 619 (1998).
[6] S. A. Schmitz *et al.*, Radiology **202**, 399 (1997).
[7] W. S. Kendal, J. Theor. Biol. **211**, 29 (2001).
[8] W. S. Kendal, BMC Cancer **5**, 138 (2006).
[9] R. Bartoszyński *et al.*, Math. Biosci. **171**, 113 (2001).
[10] L. Hanin, J. Rose, and M. Zaider, J. Theor. Biol. **243**, 407 (2006).
[11] V. P. Zhdanov, Eur. Biophys. J. **37**, 1329 (2008).
[12] K. Iwata, K. Kawasaki, and N. Shigesada, J. Theor. Biol. **203**, 177 (2000).
[13] A. K. Laird, Br. J. Cancer **19**, 278 (1965).
[14] M. Gyllenberg and G. F. Webb, J. Math. Biol. **28**, 671 (1990).
[15] N. Weidner, J. P. Semple, W. R. Welch, and J. Folkman, N. Engl. J. Med. **324**, 1 (1991).
[16] J. Folkman, Nat. Med. **1**, 27 (1995).
[17] J. W. Baish and R. K. Jain, Cancer Res. **60**, 3683 (2000).
[18] N. Shigesada and K. Kawasaki, *Biological Invasions: Theory and Practice* (Oxford University Press, Oxford, UK, 1997).
[19] D. Barbolosi, A. Benabdallah, F. Hubert, and F. Verga, Math. Biosci. **218**, 1 (2009).
[20] K. A. Moore and I. R. Lemischka, Science **311**, 1880 (2006).
[21] D. T. Scadden, Nature **441**, 1075 (2006).
[22] V. P. Zhdanov, Phys. Rev. E **79**, 061913 (2009).
[23] J. Peter and W. Semmler, IEEE Trans. Nucl. Sci. **51**, 2628 (2004).
[24] I. J. Fidler, Nat. Rev. Cancer **3**, 453 (2003).
[25] P. Waliszewski and J. Konarski, Chaos Sol. Fract. **16**, 665 (2003).
[26] R. L. Rowell and A. B. Levit, J. Colloid Interface Sci. **34**, 585 (1970).