Quantitative mathematical modeling of clinical brain metastasis dynamics in non-small cell lung cancer

M. Bilous[1,2], C. Serdjebi[3], A. Boyer[3,4], P. Tomasini[4], C. Pouypoudat[5], D. Barbolosi[3], F. Barlesi[3,4], F. Chomy[6] and S. Benzekry[1,2*]

1 : *MONC team, Inria Bordeaux Sud-Ouest, Talence, France.*

2 : *Institut de Mathématiques de Bordeaux, Bordeaux University, Talence, France.*

3 *: SMARTc Unit, Center for Research on Cancer of Marseille (CRCM), Inserm UMR 1068, CNRS UMR 7258, Aix-Marseille University U105. Marseille, France.*

4 : *CRCM, Inserm UMR 1068, CNRS UMR 7258, Aix Marseille University, Assistance Publique Hôpitaux de Marseille. Marseille. France.*

5 *: Radiation oncology department, Haut-Lévêque Hospital, Pessac, France*

6 : *Clinical oncology department, Institut Bergonié, Bordeaux, France*

[*] correspondence to Dr Sebastien Benzekry: sebastien.benzekry@inria.fr

Abstract

Brain metastases (BMs) are associated with poor prognosis in non-small cell lung cancer (NSCLC), but are only visible when large enough. Therapeutic decisions such as whole brain radiation therapy would benefit from patient-specific predictions of radiologically undetectable BMs. Here, we propose a mathematical modeling approach and use it to analyze clinical data of BM from NSCLC.

Primary tumor growth was best described by a gompertzian model for the pre-diagnosis history, followed by a tumor growth inhibition model during treatment. Growth parameters were estimated only from the size at diagnosis and histology, but predicted plausible individual estimates of the tumor age (2.1- 5.3 years). Multiple metastatic models were assessed from fitting either literature data of BM probability (n = 183 patients) or longitudinal measurements of visible BMs in two patients. Among the tested models, the one featuring dormancy was best able to describe the data. It predicted latency phases of 4.4 - 5.7 months and onset of BMs 14 - 19 months before diagnosis. This quantitative model paves the way for a computational tool of potential help during therapeutic management.

Introduction

Lung cancer is the leading cause of cancer-related death worldwide[1]. Nearly 80% of lung cancers are non-small cell lung cancer (NSCLC) and 60% of them are diagnosed at the metastatic stage[2]. Brain metastases (BMs) affect more than 20% of patients with NSCLC[3,4]. Despite recent advances in this field, BMs remain a major challenge as they are associated with poor prognosis[5–8]. In addition, BMs are responsible for disabling symptoms, with impaired patients' reported outcomes.

Several biological and clinical open questions can be formulated: 1) What was the pre-diagnosis course of a patient presenting with NSCLC? 2) For a patient developing overt BMs, when did cerebral invasion first occur? 3) Were most of the BMs spread by the primary tumor (PT) or did BMs themselves represent a significant source of secondary spread[9–11]? 4) How did growth kinetics compare between the PT and the BMs? Were there dormancy periods during the BMs history[12,13]? 5) For a patient with no or few (≤ 5) BMs at a given time point, what is the risk and extent of the occult disseminated disease in the brain (if any)?

Answers to these questions would have direct clinical utility. The clinical follow-up and planning of cerebral magnetic resonance imaging (MRI) could indeed highly benefit from individualized predictions of the probability of BM relapse or development. Moreover, the prediction of the risk and timing of BM may help clinical decision between localized (e.g. stereotactic radiotherapy), pan-cerebral

(whole-brain radiotherapy (WBRT)) and/or systemic treatment. As of today, prophylactic cranial irradiation in the management of NSCLC is not recommended despite a reduction in the risk of BMs, due to the lack of overall survival benefit and important neuro-cognitive toxicities[6,14,15]. Despite recent improvement in radiotherapy techniques, with the development of hypocampal-sparing WBRT[16], whole brain volume reduction and neurocognitive function decline were described[17]. This highlights the need for a better prediction of patients with highest risk of BM development or relapse, in order to select patients with the higher benefit/risk ratio for WBRT. Several phase III trials were conducted but no firm conclusion applicable for the entire patient population could be drawn[18,19]. The prediction of the risk of BM development and indication for WBRT to prevent brain dissemination is particularly interesting for specific populations such as *EGFR* or *ALK* positive NSCLC patients, for whose incidence of BMs is high[20–22]. This points to the need of rational tools to decide therapeutic action in a patient-specific way.

For such aims, quantitative mathematical modeling may be of considerable help, by inferring underlying processes from observed data or providing useful numerical tools in the era of personalized medicine[23,24]. However, despite numerous studies in mathematical and computational oncology, the majority of the efforts have to date focused on mathematical models at the scale of a single tumor[25]. Moreover, the comparison of the models to empirical data remains infrequent. Historically, modeling efforts in the field of metastasis were first initiated by statistical models phenomenologically describing relapse hazards[26–

[29]. In the 1970's, Liotta et al. pioneered the development of biologically-based and experimentally-validated models for all the main steps of the metastatic process[30]. Since then, only a small number of studies addressed this topic[31–40]. Of specific relevance to the current work, the Iwata model[31] introduced a size-structured population approach to capture the time development of a population of regional metastases from a hepatocellular carcinoma. It was further extended and studied from the mathematical and numerical viewpoints[41–43], and was found able to reproduce experimental *in vivo* data[37–39].

To understand biological dynamics of metastasis, animal studies allowing to capture the natural course of the disease[37,38] and the effect of therapeutic interventions in controlled environments[39] provide valuable data for quantitative analysis, although with limited clinical relevance. Experimental procedures are tedious and often only provide access to the total metastatic burden, neglecting its distribution into distinct metastatic lesions when no anatomical imaging modalities are employed[38,39].

Individual-level clinical data of metastatic development are also challenging to obtain because precise number and size of existing lesions are not routinely collected. Apart from the landmark work of Iwata et al.[31] (which only included one untreated patient with hepatocellular carcinoma), no study has yet provided a quantitative modeling analysis of longitudinal data of individual-level number and size of metastases. The purpose of the current work is to develop this

approach for BMs dynamics in NSCLC, in order to address the biological and clinical questions stated above.

Results

Pre-diagnosis natural course of lung primary tumors

We investigated two possible growth models for the natural course of lung PTs: exponential and Gompertz. Exponential growth is the simplest model expressing uncontrolled proliferation and is often well adapted to describe tumor growth kinetics during limited observation periods[44]. However, it has been shown that on longer timeframes (typically under volumes increases of 100 to 1000-fold), the specific growth rate of tumors decreases[45], which is well captured by the Gompertz model[45–47]. For calibration of the parameters, we used the data of the PT size and mean doubling time at diagnosis (see Methods). The latter was assumed to be histology-dependent and retrieved from a meta-analysis comprising a total of 388 adenocarcinomas and 377 squamous cell carcinomas[44,48] (see supplementary Table S1). For the Gompertz model (two parameters), we additionally assumed a carrying capacity of 10^{12} cells[49]. Under such assumptions, we obtained predictions of pre-diagnosis phases of 19 years versus 5.4 years in the exponential and Gompertz cases, respectively, for an adenocarcinoma with diameter 35 mm (median value from ref.[50]), see Figure S1. The first prediction seems highly unrealistic compared with previous reports estimating the age of lung primary tumors to be 3-4 years old[51], based on a different method using time to recurrence[52]. The Gompertz prediction on the other hand is rather consistent with the literature range. Moreover, the

obtained value of the cellular proliferation rate was realistic, corresponding to a length of the cell cycle $\left(=\frac{\log(2)}{\alpha_{0,p}}\right)$ of 24.4 hours[53].

We therefore concluded that the Gompertz model was best adapted to describe the pre-diagnosis natural history of the PT.

Probability of brain metastasis as a function of primary tumor size can be described by a mechanistic computational model of metastasis

PT size was reported by Mujoomdar et al.[50] as a major predictive factor of BM in NSCLC. We investigated whether our mechanistic model of BM was able to fit the data from these authors, at the population level (see Methods for the definition of the model). To do so, we assumed a statistical distribution associated with the inter-individual variability for the metastatic dissemination parameter μ (see Methods). The growth parameters were determined as above, i.e. from histology only. This approach yielded size-dependent population probabilities of BM occurrence, which could be compared to the data. The model fits were reasonably accurate, for both histological types (Figure 1). Importantly, this demonstrates the descriptive power of our model, which is based on mechanistic simulations of the underlying biological process. In this regard, this is superior to a mere fit of a biologically agnostic statistical model such as a linear regression. Notably, this result was obtained with a minimal number of parameters to describe the inter-patient heterogeneity in metastatic potential.

Namely, only two parameters defining the lognormal statistical distributions of μ in the populations. The coefficients of variation were very large (> 6,000%). The median value μ_{pop} gives a quantitative way to measure median BM aggressiveness. It was found two $\log_{10}$ orders of magnitude larger for adenocarcinomas than for squamous cell carcinomas ($\mu_{pop} = 1.39 \times 10^{-11}$ vs $\mu_{pop} = 1.76 \times 10^{-13}$).

Model-based comparison of biological scenarios with individual longitudinal data suggests dormancy

Data

These consisted of 10 and 11 PT size measurements, and 47 and 16 BM size measurements in two patients, respectively (one with an adenocarcinoma, the other with a squamous cell carcinoma). We first analyzed the adenocarcinoma patient (Figure 2) to develop the model and used the second patient to investigate the generalizability of our findings. In the adenocarcinoma patient, the PT first responded to systemic therapy before slowly regrowing (Figure 2A). A first distant BM was detected 20 months after diagnosis, and kept growing uncontrolled afterwards (Figure 2B). Other BMs appeared during follow-up (Figure 2C), reaching a total of 20 BMs at 48 months, date of last examination (Figure 2D).

Growth kinetics

To model the effect of systemic treatment on the PT, we found that a tumor growth inhibition model[54] (equation (1)) was able to adequately fit the data (Figure 2A).

Interestingly, for the BM kinetics, we found that the untreated PT Gompertz model predicted surprisingly well the observed data (Figures 2B and S2). Specifically, when employing the Gompertz parameters used for the pre-diagnosis PT history, and using initial conditions on each BM, the future growth curves were accurately predicted. This indicated that: 1) for this patient, the BMs did not respond to systemic therapy, at least from the time of first observation, 2) it is reasonable to assume that all BMs grew at the same growth rate, 3) the BMs growth rate might be similar to the PT growth rate, at least during the clinically overt phase.

Quantitative assessment of five biologically-based models of metastatic dissemination and colonization

However, this mere description of the BMs growth is not satisfactory as a model of systemic disease, since the dissemination part is absent. In particular, no model is given for the initiation times of the BMs. To include the dissemination component of the metastatic process, we relied on a modeling framework first initiated by Iwata et al.[31] (see Methods).

Five models, corresponding to five possible biological scenarios, were investigated (Figure 3). These included: (A) a “basic” model with same growth

model and parameters for the PT and the BMs and no secondary dissemination (i.e. BMs spread only by the PT); (B) a model allowing different growth parameters for the PT and the BMs; (C) a model with secondary dissemination, i.e. the ability of BMs to spread BMs themselves[31]; (D) a model allowing a delay before initiation of metastatic ability of the PT (the so-called linear model in which dissemination occurs at a late stage, opposed to the parallel model where dissemination is an early event[49]) and (E) a model allowing BM dormancy phases, i.e. time periods with null growth rate[12,13,55]. Each of these scenario was translated into corresponding mathematical equations (see Methods).

The models exhibited different descriptive power, as quantified by the best-fit value of the objective function (Table 1). The best description of the data was obtained by the dormancy model (Figure 4). Importantly, this model achieved accurate goodness-of-fit for 47 BM size measurements over 6 time points (Figure 4A-B) with only 3 parameters fitted from the BM sizes data (μ, τ and γ). Consequently, parameter identifiability was found excellent (relative standard errors < 5%, Table 2). Apparition of the first BM above the visibility threshold was also accurately described (19 vs 20 months, model vs data, Figure 4C). The model overestimated the number of small visible BMs at 40 months, possibly because a lot of these were present but too small to be visible at MRI. Indeed, a lot of BMs were present at 43 months – in accordance with the model (Figure 4A) – which thus emphasizes its predictive potential.

The basic (A), secondary (C) and delay models (D) exhibited substantially worse goodness-of-fit (Table 1 and Figures S3-5), suggesting that their underlying hypotheses are insufficient to provide valid explanations of BM dynamics. In the delay model, the fit was improved compared to models (A) and (B) but the value of the delay t_d was found unrealistically large (4.8 years after the first cancer cell, corresponding to 6 months before diagnosis).

Finally, model (D) yielded a significant improvement of the goodness-of-fit for both the dynamics of the number of visible BMs and cumulative size distributions, while not deteriorating the practical identifiability of the parameters (Figure S6). Of note, the estimated values of the BM growth parameters were close to the PT ones. However, given the previous observation that during the clinical phase the BMs grew at a growth rate consistent with the one predicted for the PT, combined with the facts that the optimal objective value was achieved for the dormancy model and that it is more parsimonious (one parameter less), we selected the dormancy model for further analysis.

Clinically relevant simulations of the disease course reveal times of BM initiation

From the quantitative calibration of the dormancy model to the data, several predictions of clinical interest can be made. The value of γ that generated the best-fit was 1, suggesting equal probability for all the cells within the PT to disseminate. In turn, this could indicate of a well-vascularized tumor, which might be prone to anti-angiogenic therapy. Interestingly, the value of μ inferred

from this patient-level data was in the same range as the one inferred from the above population analysis ($\mu = 2 \times 10^{-12}$ cell^{-1}.day^{-1} versus $\mu_{pop} = 1.39 \times 10^{-11}$ cell^{-1}.day^{-1}), giving further support to our approach.

Once calibrated from the data, our model allowed to simulate the predicted natural history of the disease. The supplementary movie S1 shows a simulation of the PT growth together with the apparition of the entire population of BMs (visible + invisible). Stars indicate tumors that are invisible (< 5 mm), and the BMs size distribution time course is also simulated. BMs represented in gray were born before the diagnosis, while BMs in white are the ones born after.

Apart from the age of the PT, prediction of birth times and invisible BM burden at any time point can be performed. Interestingly, we found that the first BM – which clinically appeared 19 months after diagnosis – was already present 14 months before clinical detection of the PT (Figure 5A). In fact, at the time of diagnosis, our model predicted the presence of 10 occult BMs, representing a total burden of 1,167 cells mostly distributed into the largest (first) BM (size 1,046 cells ≃ 0.126 mm), see Figure 5B. Notably, at the time the PT had reached the imaging detection limit, no metastasis had occurred yet. This suggests that if the disease had been detected earlier (e.g., by systematic screening), BM occurrence might have been prevented provided the tumor had been operable at this time. The amount of predicted BMs present at T_2 was 28 tumors, with an overall BM burden of 1.7×10^7 cells (Figure 5C). Therefore,

provided that neuro-cognitive risks would be acceptable, the model would recommend WBRT or systemic treatment rather than localized intervention only.

Together, these results demonstrate the potential clinical utility of the model for prediction of the invisible BM, in order to inform therapeutic decision.

Generalization to a second patient

To test whether the dormancy model was generalizable, we used data from a second patient, which was not employed during the model development phase. Given the different histological type of the lung PT (squamous cell carcinoma), we adapted the doubling time accordingly (Table S1) and found a younger age of the PT of 2.1 years. Estimation of the log-kill parameter from the TGI PT model during therapy suffered from lack of identifiability, due to an estimated short duration of treatment effectiveness (see parameter κ in Table 2). Response of the PT to therapy was indeed characterized by a faster relapse (parameter α_1), compared with the first patient. The qualitative structure of the model confirmed its descriptive power by being able to give an accurate description the BMs size dynamics (Figures S7-8), while the “basic” and “secondary dissemination” models were not (data not shown). Several parameters appeared to be patient-dependent, such as the dormancy duration τ, estimated to 171 days versus 133 days for the first patient. Due to the lower number of data points available for this patient, parameters identifiability was found worse, but still acceptable (Table 2).

Resulting clinical predictions were distinct (Figure S8), emphasizing the patient-specific nature of BM dynamics. The first BM was clinically detected 14 months after diagnosis, but was predicted to have been disseminated 45 months prior to diagnosis. While for both patients cerebral dissemination had already occurred at the time of diagnosis, its extent was different with a much larger mass (>1,000 cells) for the first patient than for the second (8 cells). This is due to the long period of dormancy for patient 2, resulting in all 8 predicted BMs being still dormant at the time of diagnosis.

These results demonstrate the flexibility of our model, which is able to describe and predict distinct situations in terms of repartition of the metastatic burden into individual lesions.

Discussion

Using both population- and individual-level data of BM development in NSCLC, we have developed a general method based on biologically grounded computational models that allowed: 1) to infer the disease age from data on PT size and histology only, 2) to test multiple scenarios of metastatic dissemination and colonization against macroscopic data available in the clinic and 3) to infer times of BM initiations and number and size of invisible lesions during the clinical course of the disease.

Estimation of the duration of the pre-diagnosis PT growth phase has important implications, since BMs are more likely to have occurred for an old PT compared

to a young one. Our results showed a significant difference whether considering exponential or Gompertz growth, which was consistent with previous findings[48]. Of note, our age estimates of 5.3 and 2.1 years-old are in relative agreement with the 3-4 years range found by others relying on a different method[51].

Importantly, we were then able to include this description of PT past growth into a mechanistic model to describe the probability of BM occurrence. We found that the cell-scale dissemination parameter μ was significantly larger in adenocarcinomas than in squamous cell carcinomas, which offers a quantitative theory for the reported differences in BM occurrence between these two histological types[50].

More generally, our computational platform provided a way to translate biological findings into clinically useful numerical tools. However, in order to provide robust inference, the complexity of the models had to remain commensurate with the available data. Thus, several higher order phenomena relevant to the metastatic process were ignored or aggregated into mesoscopic parameters. For instance, the metastatic potential μ is the product of several cell-scale probabilities relating to the multiple steps of the metastatic cascade[56]. The median values inferred from population analyses based on probabilities of BM ($\mu_{pop} = 1.39 \times 10^{-11}$ and 1.76×10^{-13}) matched relatively well the ones that were found from analysis of patient-specific data ($\mu = 2 \times 10^{-12}$ and 1.02×10^{-12}), given the standard deviations. The larger value found in patient 1 could be due to histology or to the *EGFR* mutational status, known to impact

on the BM aggressiveness[3,22,57,58]. Interestingly, while relying on distinct modeling techniques, the value of μ that we obtained was in the same range as obtained by others using stochastic evolutionary modeling[34,59].

When evaluating the descriptive power of multiple biologically-based models against longitudinal data of BM sizes, our results suggested possible phases of dormancy. This finding is supported by several preclinical and clinical reports[13,60]. For instance, Nguyen et al. mention possible dormancy phases for cerebral metastases from lung cancer[13]. Durations of these latent periods are mentioned to be of the order of months, which is consistent with our quantitative results (4.4 and 5.7 months for patients 1 and 2, respectively). BM dormancy could possibly be explained by the fact that colonizing lung carcinoma cells face a foreign micro-environment when reaching the brain, which hampers their establishment and growth[61]. This could cause single cells to remain quiescent or tumors to remain at a small, pre-angiogenic size[12,62]. Indeed, vascular endothelial growth factor-A (*VEGF*-A) has been shown to mediate BM dormancy in a mouse model of lung cancer[60].

Several second-order phenomena were ignored here for the sake of identifiability, but could nevertheless impact BM dynamics and dormancy. These include tumor-tumor interactions, either through soluble circulating factors[63] distantly inhibiting growth and possibly maintaining dormancy[62], or by exchange of tumor cells between established lesions[9,64]. We have recently proposed a

model for distant interactions that was validated in a two-tumors experimental system and could be incorporated into the current modeling platform[65,66].

While not problematic for the two patients that were investigated here since BMs did not respond to the systemic treatment – possibly because of the blood-brain barrier hampering delivery of the anti-cancer agents – a major limitation of our model is that the effect of therapy on the BMs was neglected. We intend to address this in future work, in particular for optimizing and personalizing combination therapies[24,67,68]. Moreover, the current study needs to be extended to a larger number of patients.

Our study demonstrates the potential clinical utility of our computational platform as a personalized predictive tool of BM that could be helpful in the decision of WBRT versus localized treatment only. For instance for patient 1, at the time of detection of the first BM or at 40 months when only 6 BMs were visible, large amounts of invisible BMs were predicted by the model, which would advise in favor of WBRT. In contrast, a smaller metastatic burden was predicted in patient 2, thus advising for localized intervention only. Thus, our quantitative modeling approach could complement existing predictive metrics such as the brain metastasis velocity index[69,70], for patients undergoing stereotactic radiosurgery (SRS), i.e. patients presenting with a small (≤ 3-5) number of BMs. While the latter requires data at a second time point after a first SRS, our model could predict the invisible burden of BM already at the first SRS, and consequently inform the risk and timing of future relapse.

In order to translate our findings into a clinically usable tool, further methods need to be developed to calibrate the parameters of the model from data already available at diagnosis or at the first BM occurrence. By relying on a small number (3) of free parameters for the BM dynamics, our model represents a realistic candidate for such a purpose. To this respect, in addition to routine clinico-demographic features, molecular gene expression signatures[71] (for parameter μ, for instance) as well as radiomics features[72] (for parameter γ) might represent valuable resources.

Materials and methods

Patient data

The data used in this study concerned patients with non-small cell lung cancer (NSCLC) and were of two distinct natures: 1) population data of probability of BM as a function of PT size retrieved from the literature[50] and 2) longitudinal measurements of PT and BM diameters in two patients with NSCLC retrieved from imaging data (CT scans for lung lesions, MRI for brain tumors). Both patients had unresectable PT at diagnosis. The first patient (used for model development) was extracted from an EGFR mutated cohort from Institut Bergonié (Bordeaux, France). The second patient (used for model validation) had an EGFR wild type squamous cell lung carcinoma and was issued from routinely treated patients in the thoracic oncology service of the University Hospital of Marseille. The data comprised 10 PT sizes and 47 BM sizes

(spanning 6 time points) for the first patient and 11 PT sizes and 16 BM sizes (spanning 4 time points) for second patient.

Data use from the Marseille patient was approved by a national ethics committee (CEPRO, Comité d'Evaluation des Protocoles de Recherche Observationnelle, reference number 2015-041), according to French law. Data use from the Marseille patient was approved by the institutional ethics committee of Institut Bergonié (Comité de Protection des Personnes Sud-Ouest et Outre Mer III) and IRB approval was obtained for use of the images. The need for written informed consent of data collection was waived for this patient, in accordance with the related policy of Institut Bergonié. The study was performed in accordance with the Declaration of Helsinki, Good Clinical Practices, and local ethical and legal requirements.

Mathematical modeling of primary tumor growth and metastatic development

Primary tumor growth

The pre-diagnosis natural history of the primary tumor size $S_p(t)$ for times $t < T_d$ (diagnosis time) – expressed in number of cells in the model computations – was assumed to follow the Gompertz growth model[46,47], i.e.

$$\begin{cases} \frac{dS_p}{dt} = \left(\alpha_{0,p} - \beta_p \ln(S_p)\right)S_p \\ S_p(t=0) = 1 \end{cases} \Longrightarrow S_p(t) = e^{\frac{\alpha_{0,p}}{\beta_p}\left(1-e^{-\beta_p t}\right)}, \quad \forall t < T_d$$

where time $t = 0$ corresponds to the first cancer cell, parameter $\alpha_{0,p}$ is the specific growth rate (i.e. $\frac{1}{S_p}\frac{dS_p}{dt}$) at this time and β_p is the exponential rate of decrease of the specific growth rate. Conversions from diameter measurements to number of cells were performed assuming spherical shape and the classical assumption 1 mm^3 = 10^6 cells[73]. After treatment start (T_d), the primary tumor size was assumed to follow a tumor growth inhibition model[54] consisting of: 1) exponential growth (rate α_1), 2) log-kill effect of the therapy (efficacy parameter κ)[74] and 3) exponential decrease of the treatment effect due to resistance, with half-life t_{res}. The equation is:

$$\frac{dS_p}{dt} = \alpha_1 S_p - \kappa e^{-\frac{t-T_d}{t_{res}}\ln 2} S_p, \quad \forall t \geq T_d. \tag{1}$$

Metastatic development

(A, B). Basic and different growth models

The general modeling framework we employed was derived from the work of Iwata et al.[31]. It consists in modeling the population of metastases by means of a size-structured density $\rho(t,s)$, of use to distinguish between visible and invisible tumors. Metastatic development of the disease is reduced to two main phases: dissemination and colonization[75]. The multiple steps of the metastatic cascade[56] are aggregated into a dissemination rate with expression:

$$d(S_p) = \mu S_p^{\gamma}, \quad (2)$$

which corresponds to the number of successfully born BM per unit of time. In this expression, the geometric parameter γ corresponds to the intra tumor repartition of the cells susceptible to yield metastasis and μ is the per day per cell probability for such cells to overcome all the steps of the metastatic cascade (acquisition of metastatic-specific mutations, epithelial-to-mesenchymal transition, invasion of the surrounding tissue, intravasation, survival in transit, extravasation and survival in the brain). For $\gamma = 1$ all cells in the PT have equal probability to give a BM whereas a value of $\gamma = 0$ indicates a constant pool of cells having metastatic ability (cancer stem cells). Intermediate values $0 < \gamma < 1$ can be interpreted as the geometric disposition of the metastatic-able cells, including the surface of the tumor ($\gamma = 2/3$) or a fractional dimension linked to the fractal nature of the tumor vasculature[76]. Assuming further that the growth of the metastasis follows a gompertzian growth rate

$$g(s) = (\alpha_0 - \beta \ln(s))s$$

with growth parameters α_0 and β possibly equal (basic model (A)) or distinct (different growth model (B)) compared to the PT ones, the density ρ satisfies the following transport partial differential equation[31]:

$$\begin{cases} \partial_t \rho(t,s) + \partial_s\big(\rho(t,s)g(s)\big) = 0 & t \in (0,+\infty),\ \ s \in (S_0,+\infty) \\ g(S_0)\rho(t,S_0) = d\left(S_p(t)\right) & t \in (0,+\infty) \\ \rho(0,s) = 0 & s \in (S_0,+\infty) \end{cases}, \qquad (3)$$

,

where S_0 is the size of a BM at birth (here assumed to be one cell). From the solution to this equation, the main quantity of interest for comparison to the empirical data is the number of metastasis larger than a given size s (cumulative size distribution):

$$f(t,s) = \int_s^{+\infty} \rho(t,s')ds'.$$

The total number of metastases – denoted $N(t)$ – is obtained by using $s = V_0$ above and its expression can be directly computed without solving the entire problem (3) as it is given by:

$$N(t) = \int_0^t d\left(S_p(t')\right)dt'. \qquad (4)$$

Using the method of characteristics, one can derive the following relationship between N and f :

$$f(t,s) = N(t - t(s)), \qquad (5)$$

where $t(s)$ is the time for a tumor growing at rate g to reach the size s. In the case of Gompertz growth one has:

$$t(s) = -\frac{1}{\beta}\ln\left(1 - \frac{\beta}{\alpha_0}\ln(s)\right).$$

Of particular interest is the number of visible BMs $f(t, S_{vis})$ with S_{vis} the minimal visible size at CT scan taken here to be 5 mm in diameter.

(C). Secondary dissemination

In the previous model formulation, all BMs were assumed to have been seeded by the primary tumor. When BMs are also allowed to spread metastases themselves, this results in a second term in the boundary condition of (3) and the model becomes[31]:

$$\begin{cases} \partial_t \rho(t,s) + \partial_s\big(\rho(t,s)g(s)\big) = 0 & t \in (0,+\infty),\ \ s \in (S_0,+\infty) \\ g(S_0)\rho(t,S_0) = d\left(S_p(t)\right) + \int_{S_0}^{+\infty} d(s)\rho(t,s)ds & t \in (0,+\infty) \\ \rho(0,s) = 0 & s \in (S_0,+\infty) \end{cases} \quad (1)$$

In this case, formula (4) is not valid anymore, which complicates substantially the computation of the cumulative size distribution. A dedicated scheme based on the method of characteristics was employed[43].

(D). Delay model

Consideration of a delay t_0 before onset of metastatic dissemination in the model can be taken into account by remarking that

$$N_{delay}(t) = \int_{t_0}^{t} d(S_p(t')dt' = N(t) - N(t_0).$$

Thus

$$f_{delay}(t,s) = N_{delay}\big(t - t(s)\big) = N\big(t - t(s)\big) - N(t_0).$$

(E). Dormancy model

For inclusion of dormancy in the model – defined as a period of duration τ during which a newborn metastasis remains at size S_0 – the time to reach any given size $s > S_0$ becomes $t_{dorm}(s) = t(s) + \tau$. The cumulative size distribution is then given by:

$$f_{dorm}(t,s) = N\big(t - t_{dorm}(s)\big) = N(t - t(s) - \tau).$$

Discrete versions of the models

While continuous versions of the models were used for fitting the model to the data because they allow computations to be tractable, discrete versions were implemented for forward calculations, because of the small number of BMs. Briefly, in a discrete simulation, the appearance time of the i-th BM T_i was defined by

$$T_i = \inf\{t > 0; N(t) \geq i\},$$

with $N(t)$ the number of metastases as defined above. The size of the i-th BM $S_i(t)$ was then defined, for $t > T_i$, by:

$$\begin{cases} \dfrac{dS_i}{dt} = g(S_i) \\ S_i(t = T_i) = S_0 \end{cases}.$$

This discrete version corresponds to a Poisson process where the number of events (metastatic births) $\mathcal{N}(t)$ is replaced by its expectation: $N(t) = \mathbb{E}[\mathcal{N}(t)]$. For further links between stochastic (Poisson process) and continuous versions of the Iwata model, the reader is referred to[77]. Simulations reported in Figures 5B-C and 6 were generated using discrete versions of the models.

Models fitting and estimation of the parameters

Pre-clinical primary tumor growth

To parameterize the Gompertz function defining the PT growth, two parameters needed to be defined ($\alpha_{0,p}$and β_p). In the absence of longitudinal measurements of the PT size without treatment, these two parameters were determined from two considerations: 1) the maximal reachable size (carrying capacity, equal to $e^{\frac{\alpha_{0,p}}{\beta_p}}$) of a human tumor is 10^{12} cells[29,49] and 2) the histology-dependent value of the doubling time at diagnosis, retrieved from a meta-analysis of published literature about the natural growth of lung PTs (see Table S1, extended from[44,48]). The latter yielded values of 201 days for an adenocarcinoma and 104 days for an undifferentiated carcinoma. For the Gompertz model, the doubling time is size-dependent and its value for the PT diagnosis size $DT(S_d)$ is given by:

$$DT(S_d) = -\frac{1}{\beta_p}\ln\left(\beta_p \frac{\ln(2S_d)}{\beta_p \ln S_d - \alpha_{0,p}}\right).$$

Using the formula linking $\alpha_{0,p}$ and β to the carrying capacity, this nonlinear equation was numerically solved.

Population level : probability of BM apparition

To fit the data from[50] describing the probability of BM in a population of lung adenocarcinoma patients, we employed a previously described methodology[39,78]. Briefly, we considered that the probability of developing BM after diagnosis was the probability of having already BM at diagnosis, i.e. $\mathbb{P}(N(T_d) > 1)$, with N defined by equation (4). Of note, since this quantity only depends on the dissemination rate, this probability is equal for all the model scenarios introduced above. We fixed the value of the PT growth parameters as described above from the cohort histology and set $\gamma = 1$ as the simplest dissemination model. The inter-individual variability was then minimally modeled as resulting from a lognormal population distribution of the parameter μ $\left(\ln \mu \sim \mathcal{N}(\ln(\mu_{pop}), \mu_\sigma)\right)$. Uniform distributions of the PT diameters were assuming within each interval $(S^i, S^{i+1})$of the PT sizes $S^1, \dots, S^6$ given as data. The probability of developing a metastasis with a PT size $S \in (S^i, S^{i+1})$ writes:

$$p^i(\mu_{pop}, \mu_\sigma) = \mathbb{P}\left(\mu \int_0^{T_D(S)} S_p(t)dt > 1\right).$$

The best-fit of these probabilities – evaluated by Monte Carlo simulations – to the empirical data was then determined by least squares minimization performed using the function *fminsearch* of Matlab (Nelder-Mead algorithm)[79].

Individual level : description of longitudinal data of number and size of BM

growth

Maximum likelihood estimation

Due to the discrete nature of the data at the individual level (diameters of a small number of BMs at discrete time points), a direct comparison between the size distribution ρ solution of the problem (3) was not possible. Instead, we compared the data to the model by means of the cumulative size distribution. Denoting by t_j the observation times, x_j^i the sorted BM sizes at time t_j and y_j^i the number of metastases larger than size x_j^i at time t_j, we considered the following nonlinear regression problem:

$$y_j^i = f\left(t_j, x_j^i; \theta\right) + \sigma\varepsilon_j^i, \quad \varepsilon_j^i \sim \mathcal{N}(0,1),$$

where $\theta = (\alpha_{0,p}, \beta_p, \alpha_0, \beta, \mu, \gamma, t_0, \tau)$ regroups all the parameters of the model. Note that, except for the "secondary dissemination" model, all models can be viewed as submodels of a general model including all these parameters (the "basic model" consisting of the case $\alpha_{0,p} = \alpha_0, \beta_p = \beta, t_0 = 0$ and $\tau = 0$, for instance). Classical maximum likelihood estimation then leads to the following estimate:

$$SSE(\theta) = \sum_{i,j} \left(y_j^i - f\left(t_j, x_j^i; \theta\right)\right)^2 \tag{2}$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, l(\theta).$$

Parameters identifiability

Standard errors can be computed from this statistics' covariance matrix, given by[80]:

$$Cov = \hat{\sigma}\left(J(\hat{\theta})^T J(\hat{\theta})\right)^{-1},$$

where $J(\hat{\theta})$ is the jacobian matrix of the model (with respect to the parameter vector θ) at all time points and all sizes, evaluated at the optimal parameter $\hat{\theta}$ and $\hat{\sigma}$ is the a posteriori estimate of σ given by

$$\hat{\sigma} = \frac{1}{N-P} l(\hat{\theta})$$

with N the total number of data points and P the number of free parameters.

Using the standard errors as an identifiability metric, we repeatedly observed a lack of identifiability of parameters μ and γ when fitted together. Indeed, standard errors for μ and γ were larger than 200% when fitting the basic model to the data. Further investigation of the shape of the objective function confirmed this lack of idenfiability (Figure S7). To address this issue, we only considered a finite set of relevant possible values for γ and only optimized the value of μ. These values were $\left(0, 0.4, 0.5, \frac{2}{3}, 1\right)$ and corresponding initial conditions for μ were $(10^{-3}, 10^{-5}, 10^{-8}, 10^{-9}, 10^{-12})$. When more parameters were let free in the model (delay t_d or dormancy period τ), we generated 4×4 parameters grids for initial conditions (with $t_d \in \{0,500,1700,2000\}$ or $\tau \in \{0, 50, 100, 350\}$). Of these

16 optimization problems, the one with the minimal value of l at convergence was selected.

References

1. Ferlay, J. *et al.* GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. (2013).

2. Noone, A. *et al.* SEER Cancer Statistics Review, 1975-2015, National Cancer Institute. *National Cancer Institute* (2018).

3. Tomasini, P. *et al.* EGFR and KRAS Mutations Predict the Incidence and Outcome of Brain Metastases in Non-Small Cell Lung Cancer. *Int J Mol Sci* **17**, 2132 (2016).

4. Barnholtz-Sloan, J. S. *et al.* Incidence proportions of brain metastases in patients diagnosed (1973 to 2001) in the Metropolitan Detroit Cancer Surveillance System. *J Clin Oncol* **22**, 2865–2872 (2004).

5. Oh, Y. *et al.* Number of metastatic sites is a strong predictor of survival in patients with nonsmall cell lung cancer with or without brain metastases. *Cancer* **115**, 2930–2938 (2009).

6. Barlesi, F. *et al.* Management of brain metastases for lung cancer patients. *Bull Cancer* **100**, 303–308 (2013).

7. Métellus, P. *et al.* Place of surgery in brain metastases. *Bull Cancer* **100**, 51–56 (2013).

8. Tabouret, E. *et al.* Surgical resection of brain metastases from breast cancer in the modern era: clinical outcome and prognostic factors. *Anticancer Res* **33**, 2159–2167 (2013).

9. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357 (2015).

10. Weiss, L. Patterns of Metastasis. *Cancer Metastasis Rev.* **19**, 281–301 (2000).

11. Sugarbaker, E. V., Cohen, a M. & Ketcham, a S. Do metastases metastasize? *Ann Surg* **174**, 161–6 (1971).

12. Chambers, A. F., Groom, A. C. & MacDonald, I. C. Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer* **2**, 563–572 (2002).

13. Nguyen, D. X., Bos, P. D. & Massagué, J. Metastasis: from dissemination to organ-specific colonization. *Nat Rev Cancer* **9**, 274–284 (2009).

14. Snider, J. W., Gondi, V., Brown, P. D., Tome, W. & Mehta, M. P. Prophylactic cranial irradiation: recent outcomes and innovations. *CNS Oncol* **3**, 219–230 (2014).

15. Tallet, A. V. *et al.* Neurocognitive function impairment after whole brain radiotherapy for brain metastases: actual assessment. *Radiat Oncol* **7**, 77 (2012).

16. Gondi, V. *et al.* NRG Oncology CC001: A phase III trial of hippocampal avoidance (HA) in addition to whole-brain radiotherapy (WBRT) plus memantine to preserve neurocognitive function (NCF) in patients with brain metastases (BM). *J Clin Oncol* **37**, 2009–2009 (2019).

17. Gui, C. *et al.* A prospective evaluation of whole brain volume loss and neurocognitive decline following hippocampal-sparing prophylactic cranial irradiation for limited-stage small-cell lung cancer. *J. Neurooncol.* (2019). doi:10.1007/s11060-019-03235-7

18. Mulvenna, P. *et al.* Dexamethasone and supportive care with or without whole brain radiotherapy in treating patients with non-small cell lung cancer with brain metastases unsuitable for resection or stereotactic radiotherapy (QUARTZ): results from a phase 3, non-inferiority, randomised trial. *The Lancet* **388**, 2004–2014 (2016).

19. Pechoux, C. L., Dhermain, F. & Besse, B. Whole brain radiotherapy in patients with NSCLC and brain metastases. *The Lancet* **388**, 1960–1962 (2016).

20. Soffietti, R. *et al.* Diagnosis and treatment of brain metastases from solid tumors: guidelines from the European Association of Neuro-Oncology (EANO). *Neuro Oncol* **19**, 162–174 (2017).

21. Doherty, M. K. *et al.* Treatment options for patients with brain metastases from EGFR / ALK -driven lung cancer. *Radiother Oncol* **123**, 195–202 (2017).

22. Jiang, T. *et al.* EGFR TKIs plus WBRT Demonstrated No Survival Benefit Other Than That of TKIs Alone in Patients with NSCLC and EGFR Mutation and Brain Metastases. *J Thorac Oncol* **11**, 1718–1728 (2016).

23. Altrock, P. M., Liu, L. L. & Michor, F. The mathematics of cancer: integrating quantitative models. *Nat Rev Cancer* **15**, 730–745 (2015).

24. Barbolosi, D., Ciccolini, J., Lacarelle, B., Barlési, F. & André, N. Computational oncology - mathematical modelling of drug regimens for precision medicine. *Nat Rev Clin Oncol* (2015).

25. Byrne, H. M. Dissecting cancer through mathematics: from the cell to the animal model. *Nat Rev Cancer* **10**, 221–230 (2010).

26. Bross, D. J. & Blumenson, L. E. Statistical testing of a deep mathematical model for human breast cancer. *J Chronic Dis* **21**, 493–506 (1968).

27. Guiguet, M., Tubiana, M. & Valleron, A. J. [Size distribution of metastases during detection and adjuvant treatment: biomathematical approach]. *C R Seances Acad Sci III* **294**, 15–18 (1982).

28. Koscielny, S., Tubiana, M. & Valleron, A. J. A simulation model of the natural history of human breast cancer. *Br J Cancer* **52**, 515–524 (1985).

29. Retsky, M. W. *et al.* Computer simulation of a breast cancer metastasis model. *Breast Cancer Res Treat* **45**, 193–202 (1997).

30. Liotta, L. A., Saidel, G. M. & Kleinerman, J. Stochastic model of metastases formation. *Biometrics* **32**, 535–550 (1976).

31. Iwata, K., Kawasaki, K. & Shigesada, N. A Dynamical Model for the Growth and Size Distribution of Multiple Metastatic Tumors. *J Theor Biol* **203**, 177–186 (2000).

32. Hanin, L., Rose, J. & Zaider, M. A stochastic model for the sizes of detectable metastases. *J Theor Biol* **243**, 407–417 (2006).

33. Michor, F., Nowak, M. A. & Iwasa, Y. Stochastic dynamics of metastasis

formation. *J Theor Biol* **240**, 521–530 (2006).

34. Haeno, H. *et al.* Computational Modeling of Pancreatic Cancer Reveals Kinetics of Metastasis Suggesting Optimum Treatment Strategies. *Cell* **148**, 362–375 (2012).

35. Newton, P. K. *et al.* A Stochastic Markov Chain Model to Describe Lung Cancer Growth and Metastasis. *PLoS One* **7**, e34637 (2012).

36. Scott, J. G., Gerlee, P., Basanta, D. & Fletcher, A. G. Mathematical modeling of the metastatic process. in *Experimental Metastasis: Modeling and Analysis* (ed. Malek, A.) (Springer Netherlands, 2013).

37. Hartung, N. *et al.* Mathematical modeling of tumor growth and metastatic spreading: validation in tumor-bearing mice. *Cancer Res* **74**, 6397–6407 (2014).

38. Baratchart, E. *et al.* Computational Modelling of Metastasis Development in Renal Cell Carcinoma. *PLoS Comput Biol* **11**, e1004626 (2015).

39. Benzekry, S. *et al.* Modeling Spontaneous Metastasis following Surgery: An In Vivo-In Silico Approach. *Cancer Res* **76**, 535–547 (2016).

40. Poleszczuk, J. T. *et al.* Abscopal Benefits of Localized Radiotherapy Depend on Activated T-cell Trafficking and Distribution between Metastatic Lesions. *Cancer Res* **76**, 1009–1018 (2016).

41. Barbolosi, D., Benabdallah, A., Hubert, F. & Verga, F. Mathematical and numerical analysis for a model of growing metastatic tumors. *Math Biosci* **218**, 1–14 (2009).

42. Benzekry, S. Mathematical analysis of a two-dimensional population model of metastatic growth including angiogenesis. *J Evol Equ* **11**, 187–213 (2011).

43. Benzekry, S. Mathematical and numerical analysis of a model for anti-angiogenic therapy in metastatic cancers. *ESAIM, Math Model Numer Anal* **46**, 207–237 (2012).

44. Friberg, S. & Mattson, S. On the growth rates of human malignant tumors:

implications for medical decision making. *J Surg Oncol* **65**, 284–297 (1997).

45. Laird, A. K. Dynamics of Tumour Growth: Comparison of Growth Rates and Extrapolation of Growth Curve To One Cell. *Br J Cancer* **19**, 278–91 (1965).

46. Norton, L. A Gompertzian model of human breast cancer growth. *Cancer Res* **48**, 7067–7071 (1988).

47. Benzekry, S. *et al.* Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Comput Biol* **10**, e1003800 (2014).

48. Detterbeck, F. C. & Gibson, C. J. Turning gray: the natural history of lung cancer over time. *J Thorac Oncol* **3**, 781–792 (2008).

49. Klein, C. A. Parallel progression of primary tumours and metastases. *Nat Rev Cancer* **9**, 302–312 (2009).

50. Mujoomdar, A. *et al.* Clinical Predictors of Metastatic Disease to the Brain from Non–Small Cell Lung Carcinoma: Primary Tumor Size, Cell Type, and Lymph Node Metastases1. *Radiology* **242**, 882–888 (2007).

51. Patrone, M. V., Hubbs, J. L., Bailey, J. E. & Marks, L. B. How long have I had my cancer, doctor? Estimating tumor age via Collins' law. *Oncology (Williston Park, N.Y.)* **25**, 38-43–46 (2011).

52. Collins, V. P., Loeffler, R. K. & Tivey, H. Observations on growth rates of human tumors. *Am J Roentgenol Radium Ther Nucl Med* **76**, (1956).

53. Norton, L. & Gilewski, T. A. Cytokinetics. in *Holland-Frei Cancer Medicine.* (PMPH-USA, 2010).

54. Claret, L. *et al.* Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J Clin Oncol* **27**, 4103–4108 (2009).

55. Aguirre-Ghiso, J. A. How dormant cancer persists and reawakens. *Science* **361**, 1314–1315 (2018).

56. Gupta, G. P. & Massagué, J. Cancer metastasis: building a framework. *Cell*

127, 679–695 (2006).

57. Wang, B.-X. *et al.* Impacts of EGFR mutation and EGFR-TKIs on incidence of brain metastases in advanced non-squamous NSCLC. *Clin Neurol Neurosurg* **160**, 96–100 (2017).

58. Han, G. *et al.* A retrospective analysis in patients with EGFR-mutant lung adenocarcinoma: is EGFR mutation associated with a higher incidence of brain metastasis? *Oncotarget* **7**, 56998–57010 (2016).

59. Reiter, J. G. *et al.* Minimal functional driver gene heterogeneity among untreated metastases. *Science* **361**, 1033–1037 (2018).

60. Kienast, Y. *et al.* Real-time imaging reveals the single steps of brain metastasis formation. *Nat Med* **16**, 116–122 (2010).

61. Winkler, F. The brain metastatic niche. *J Mol Med* **93**, 1213–1220 (2015).

62. Holmgren, L., O'Reilly, M. S. & Folkman, J. Dormancy of micrometastases: balanced proliferation and apoptosis in the presence of angiogenesis suppression. *Nat Med* **1**, 149–153 (1995).

63. Dewys, W. D. A Quantitative Model for the Study of the Growth and Treatment of a Tumor and Its Metastases with Correlation between Proliferative State and Sensitivity to Cyclophosphamide. *Cancer Res* 367–373 (1972).

64. Norton, L. & Massagué, J. Is cancer a disease of self-seeding? *Nat Med* **12**, 875–878 (2006).

65. Benzekry, S., Lamont, C., Barbolosi, D., Hlatky, L. & Hahnfeldt, P. Mathematical Modeling of Tumor-Tumor Distant Interactions Supports a Systemic Control of Tumor Growth. *Cancer Res* **77**, 5183–5193 (2017).

66. Benzekry, S., Gandolfi, A. & Hahnfeldt, P. Global Dormancy of Metastases Due to Systemic Inhibition of Angiogenesis. *PLoS One* **9**, e84249-11 (2014).

67. Imbs, D.-C. *et al.* Revisiting Bevacizumab + Cytotoxics Scheduling Using Mathematical Modeling: Proof of Concept Study in Experimental Non-Small Cell

Lung Carcinoma. *CPT Pharmacometrics Syst Pharmacol* **7**, 42–50 (2018).

68. Serre, R. *et al.* Mathematical Modeling of Cancer Immunotherapy and Its Synergy with Radiotherapy. *Cancer Res* **76**, 4931–4940 (2016).

69. Farris, M. *et al.* Brain Metastasis Velocity: A Novel Prognostic Metric Predictive of Overall Survival and Freedom From Whole-Brain Radiation Therapy After Distant Brain Failure Following Upfront Radiosurgery Alone. *Int. J. Radiat. Oncol. Biol. Phys.* **98**, 131–141 (2017).

70. Yamamoto, M. *et al.* Validity of a Recently Proposed Prognostic Grading Index, Brain Metastasis Velocity, for Patients With Brain Metastasis Undergoing Multiple Radiosurgical Procedures. *Int. J. Radiat. Oncol. Biol. Phys.* **103**, 631–637 (2019).

71. Shedden, K. *et al.* Gene expression–based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* **14**, 822–827 (2008).

72. Aerts, H. J. W. L. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* **5**, 4006 (2014).

73. Spratt, J. S., Meyer, J. S. & Spratt, J. A. Rates of growth of human solid neoplasms: Part I. *J Surg Oncol* **60**, 137–146 (1995).

74. Benzekry, S. *et al.* Metronomic reloaded: Theoretical models bringing chemotherapy into the era of precision medicine. *Semin Cancer Biol* **35**, 53–61 (2015).

75. Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* **331**, 1559–1564 (2011).

76. Baish, J. W. & Jain, R. K. Cancer, angiogenesis and fractals. *Nat Med* **4**, 984–984 (1998).

77. Gomez, C. & Hartung, N. Stochastic and Deterministic Models for the Metastatic Emission Process: Formalisms and Crosslinks. in *Cancer Systems Biology* 193–224 (Humana Press, New York, NY, 2018). doi:10.1007/978-1-4939-7493-1_10

78. Barbolosi, D. *et al.* Modélisation du risque d'évolution métastatique chez les

patients supposés avoir une maladie localisée. *Oncologie* **13**, 528–533 (2011).

79. Mathworks, T. *Matlab with statistics and optimization toolboxes*. (The MathWorks, Inc., 2015).

80. Seber, G. A. & Wild, C. J. *Nonlinear regression*. (Wiley-Interscience, 2003).

Acknowledgments

The authors would like to thank Pr A. Iliadis for valuable comments and advice.

Author contributions

SB and DB conceived the research idea. CS, AB, PT, FB and FC collected the data. CP measured tumor sizes. MB and SB performed the data analysis. MB and SB programmed the software for estimation of the parameters and simulation. The paper was written by SB with editorial input from all authors.

The authors declare no potential conflicts of interest

Data availability statement

The datasets analyzed during the current study are available from the corresponding author on reasonable request

Figures and Tables

Figure 1: Probability of brain metastasis as a function of the primary tumor size

Figure 2: Data of primary tumor and brain metastases in a patient with non-small cell lung cancer

Figure 3: Schematic of the models

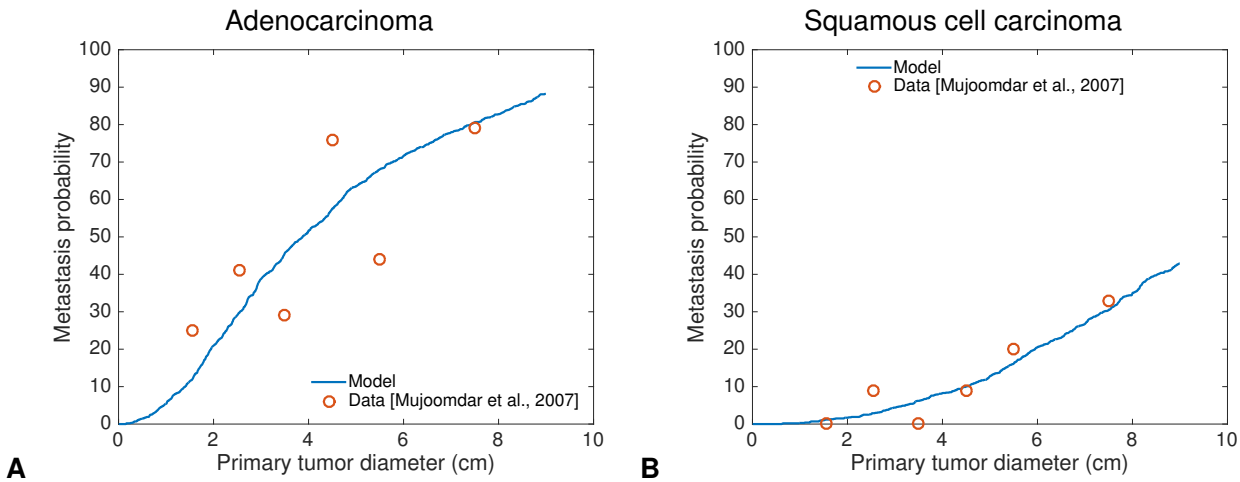
Figure 4: Fit of the dormancy model

Figure 5: Model simulated predictions of the pre- and post-diagnosis time course of the primary and cerebral disease

Table 1: Minimal value of the objective function obtained when fitting each of the models to the data

Table 2: Patient-specific parameters

Figure 1: Probability of brain metastasis as a function of the primary tumor size

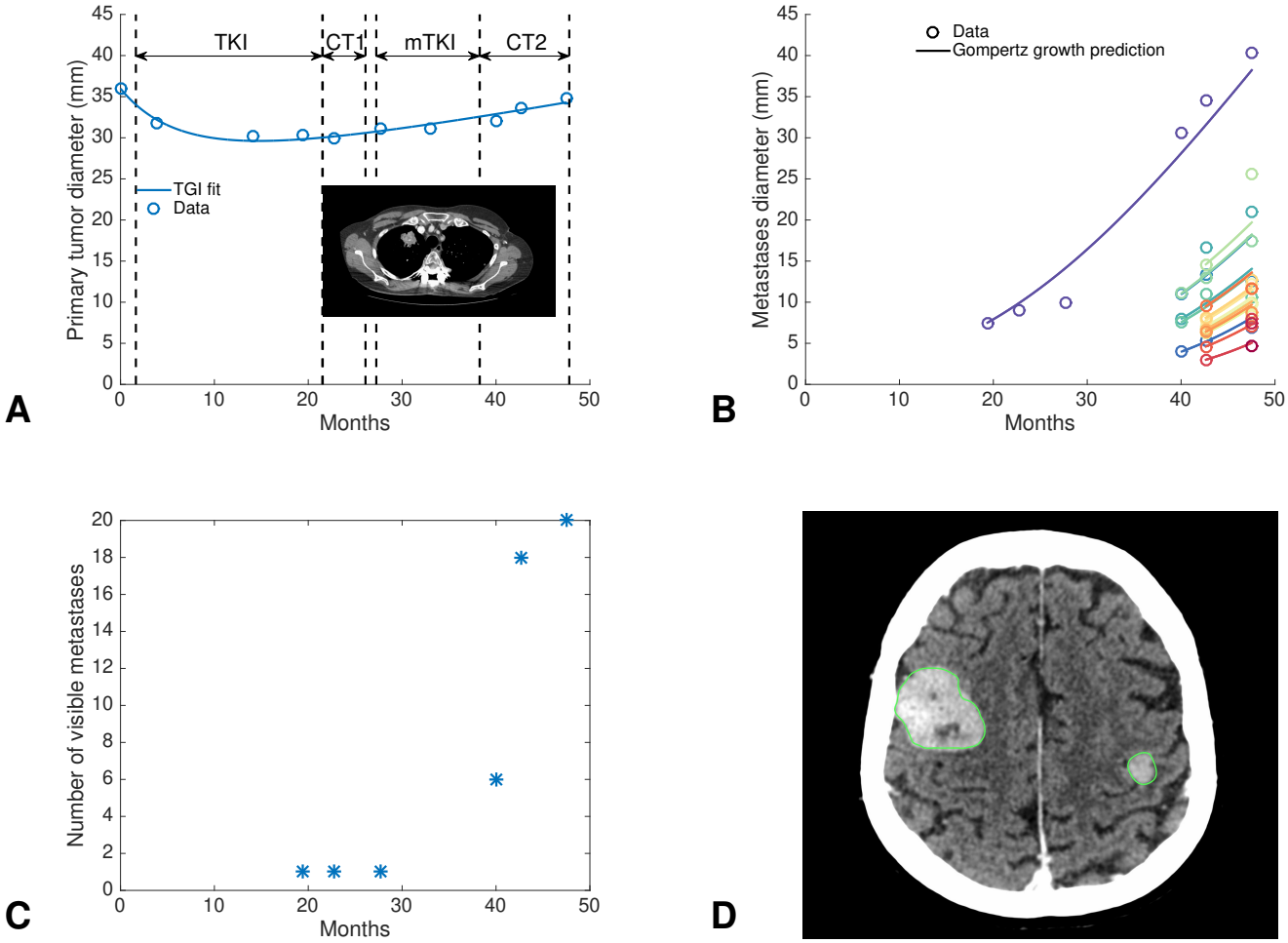


Fit of the mechanistic model for probability of BM to data from Mujoomdar et al.

A. Adenocarcinoma data (n=136). Inferred values for the distribution of μ: $\mu_{pop} = 1.39 \times 10^{-11}$ cell$^{-1}\cdot$ day^{-1} and $\mu_\sigma = 1.24 \times 10^{-9}$ cell$^{-1}\cdot$ day^{-1}

B. Squamous cell carcinoma data (n=47). Inferred values for the distribution of μ: $\mu_{pop} = 1.76 \times 10^{-13}$ cell$^{-1}\cdot$ day^{-1} and $\mu_\sigma = 1.11 \times 10^{-11}$ cell$^{-1}\cdot$ day^{-1}

Figure 2: Data of primary tumor and brain metastases in a patient with non-small cell lung cancer



A. Post-diagnosis kinetics of the primary tumor largest diameter, measured on follow-up computerized tomography images (inlet). This EGFR mutated patient was treated first with a tyrosine kinase inhibitor (erlotinib) then with several rounds of additional systemic treatments upon relapse (cytotoxic chemotherapy (docetaxel), re-challenge with erlotinib and a second cytotoxic agent (gemcitabine)), as indicated by arrows and dashed lines in the figure. The solid line corresponds to the model fit during treatment (tumor growth inhibition (TGI) model).

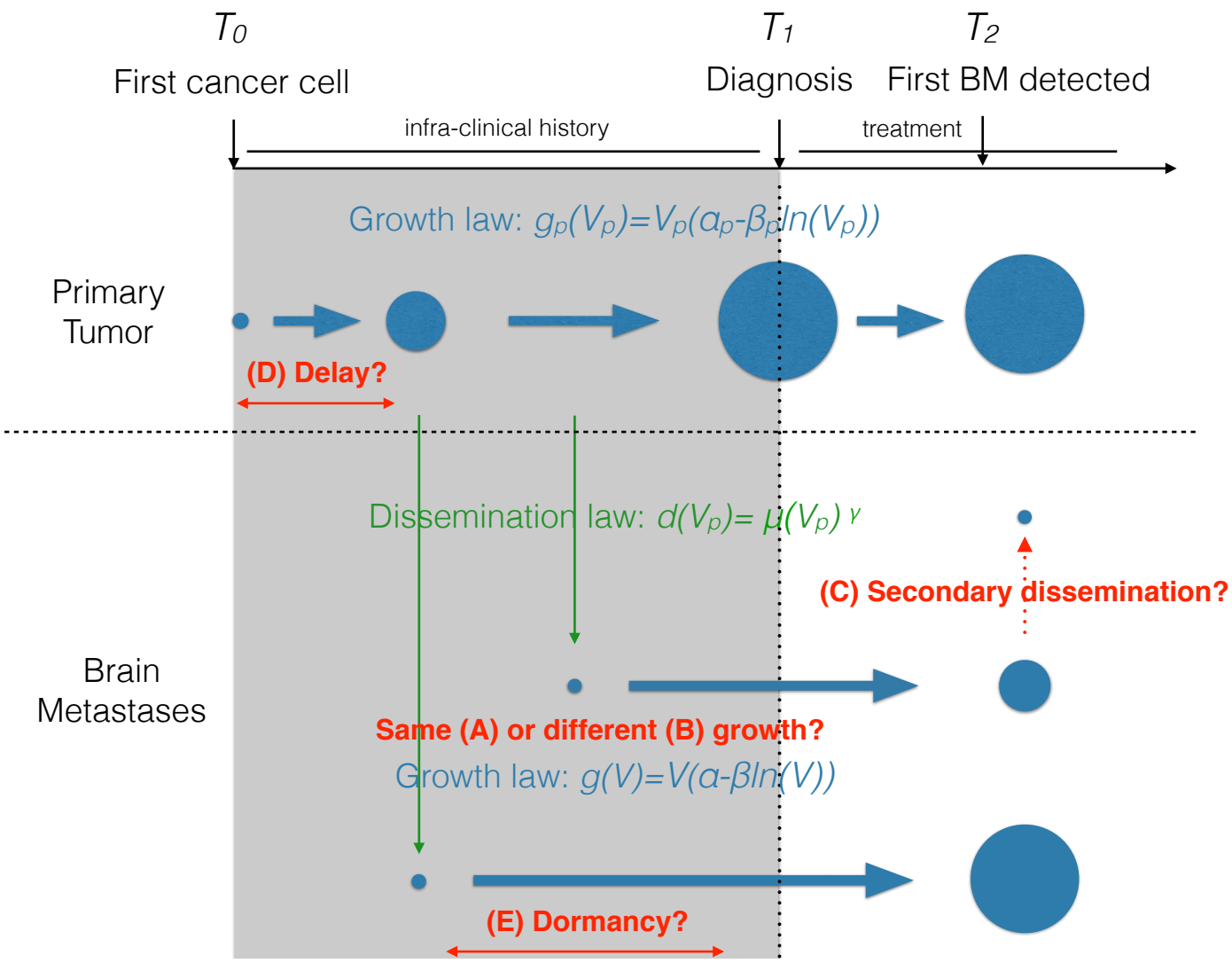
B. Growth kinetics of the brain metastases. The solid line corresponds to Gompertz growth predictions based on parameters estimated from the primary tumor size at diagnosis and histological type, using as initial condition the first observation of each BM.

C. Time course of the apparition of visible metastases.

D. Cerebral magnetic resonance image showing two brain metastases 48 months post-diagnosis (other brain lesions not visible on this slice).

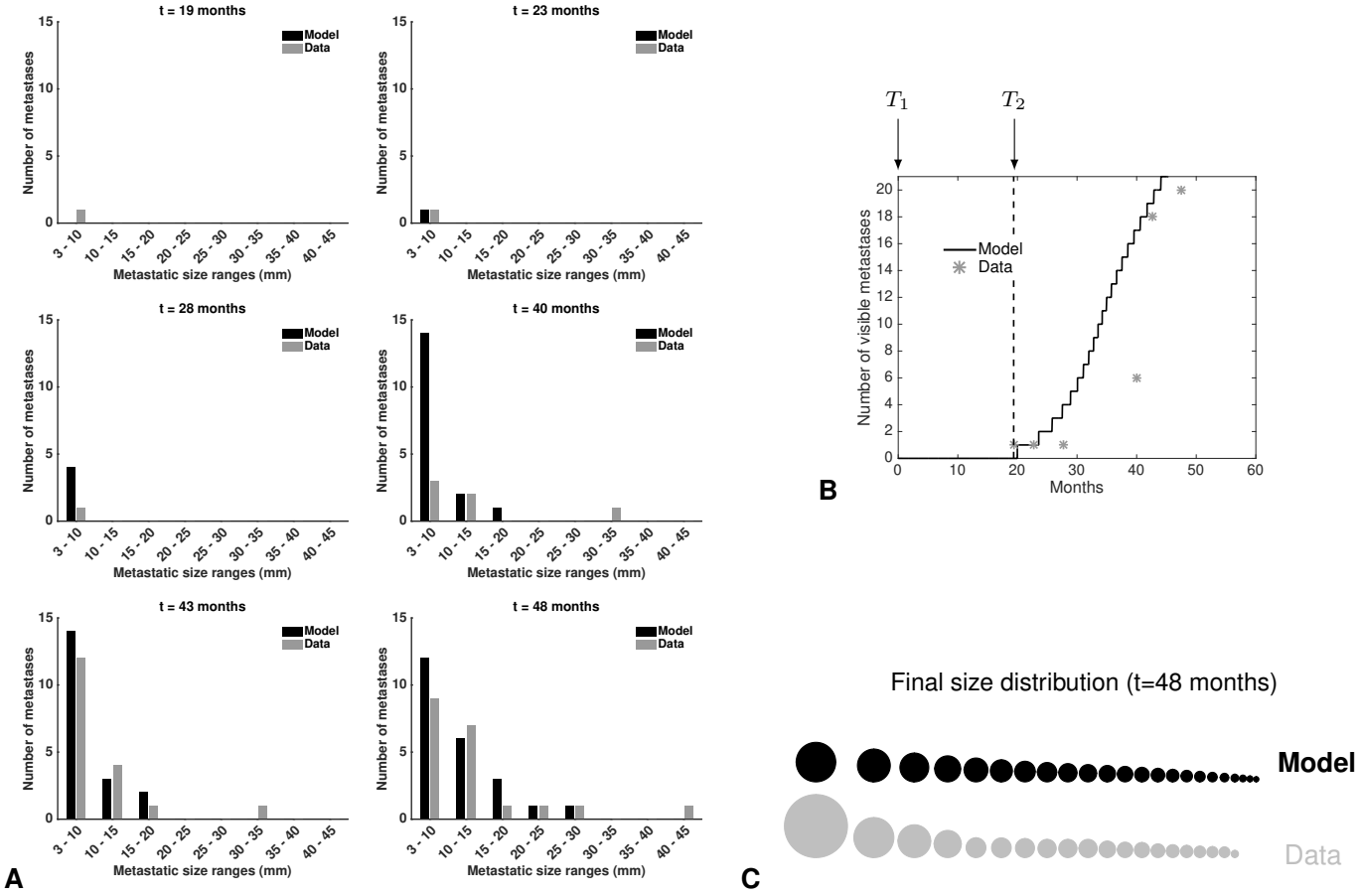
EGFR = Epithelial Growth Factor Receptor, TKI = Tyrosine Kinase Inhibitor, mTKI = maintenance Tyrosine Kinase Inhibitor, CT = (cytotoxic) chemotherapy. Time is in months from diagnosis.

Figure 3: Schematic of the models



Several biological hypotheses about metastatic dynamics were formalized as mathematical models and tested against the data. These included: same (A) or different (B) growth rates between primary and secondary tumors; only primary dissemination from the primary tumor or (C) secondary dissemination from brain metastases themselves; (D) possibility of a delay between cancer initiation and onset of metastatic ability; and (E) the possibility of a dormancy lag time between metastatic spread and growth initiation.

Figure 4: Fit of the dormancy model



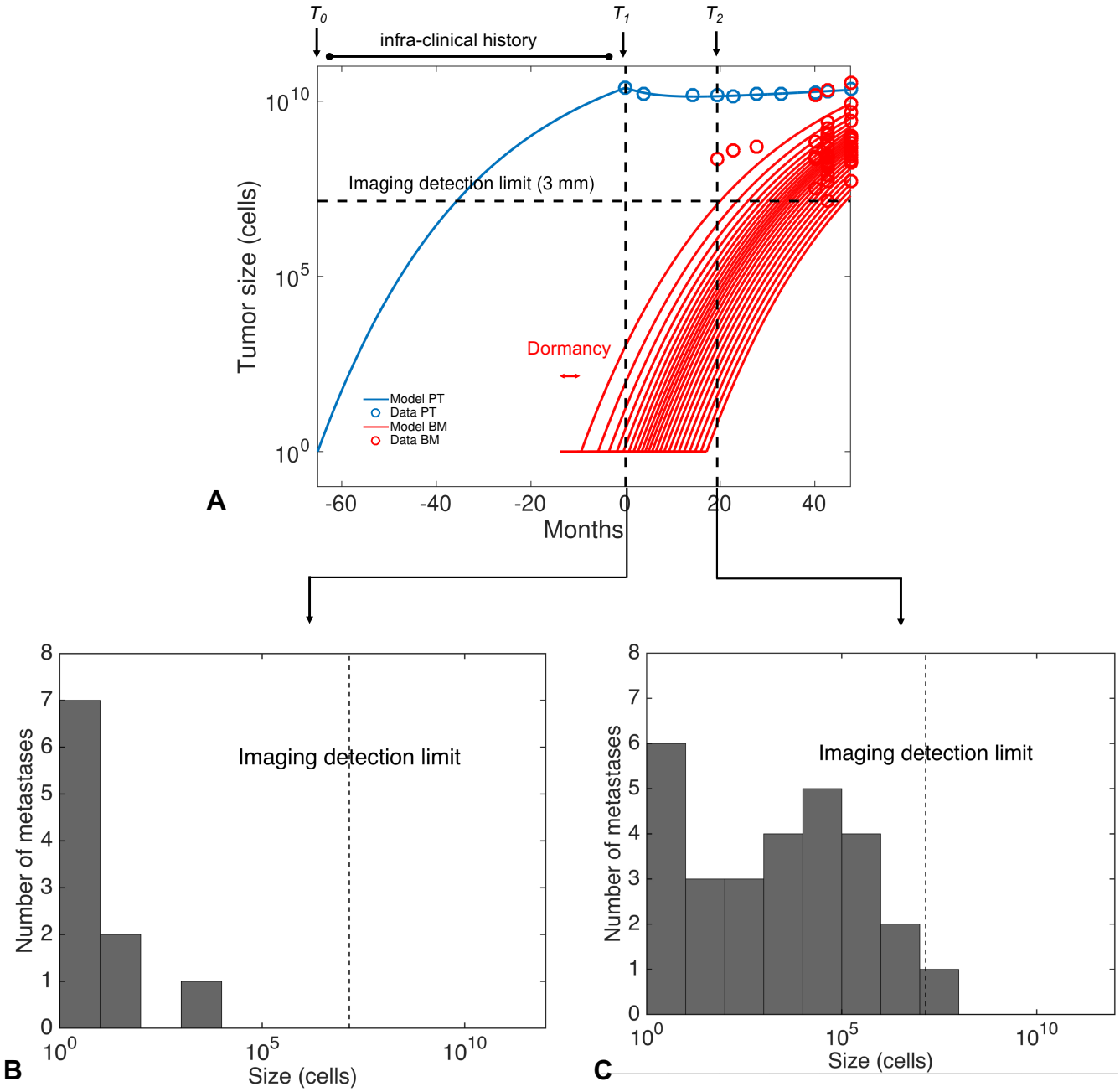
A. Time course of the visible brain metastases (BMs) size distributions during follow-up. Comparison between model calibration and data.

B. Time course of the number of visible BMs. T1 = time of diagnosis. T2 = detection time of the first brain metastasis.

C. Comparison of the BM size distribution between the model fit and the data at last follow-up (48 months post-diagnosis).

Time is in months from diagnosis.

Figure 5: Model simulated predictions of the pre- and post-diagnosis time course of the primary and cerebral disease



A. Model-inferred growth kinetics of the primary tumor (in blue) and the brain metastases (in red), compared with data (circles). T0 = time of first cancer cell,T1 = time of diagnosis, T2 = detection time of the first brain metastasis. Only brain metastases that will become visible are shown.

B. Predicted size distribution of the brain metastases at diagnosis.

C. Predicted size distribution of the brain metastases at the time of clinical occurrence of the first one.

Simulations performed with the discrete version of the model. PT = primary tumor, BM = brain metastasis. Time is in months from diagnosis.

Table 1: Minimal value of the objective function obtained when fitting each of the models to the data

Model	Patient 1	Patient 2
Basic	5.51	2.53
Secondary	5.43	2.3
Delay	5.23	1.53
Dormancy	4.93	1.71
Diff. growth	4.95	1.79

The objective function was the sum of squared residuals (see expression (7)).

Table 2: Patient-specific parameters

Parameter	Meaning	Unit	Patient 1	Patient 2	Determination
S_d	PT size at diagnosis	mm	36.0	53.7	Data
Histology			Adenocarcinoma	Squamous cell carcinoma	Data
T_d	Age of the PT at diagnosis	year	5.3	2.1	Predicted
Visibility threshold		mm	3	3	Fixed
$S_{0,p}$	Initial size of the PT	cell	1	1	Fixed
S_0	Initial size of a BM	cell	1	1	Fixed
$\alpha_{0,p}$	Proliferation rate at one cell (PT)	day^{-1}	0.0284	0.0858	Computed
β_p	Exponential decrease of growth rate (PT)	day^{-1}	1.03×10^{-3}	3.10×10^{-3}	Computed
DT_{PT}	PT doubling time at 3 cm	day	173	57	Predicted
α_1	PT growth rate during relapse	day^{-1}	5.72×10^{-4} (23.6)	0.0101 (5.28×10^{-3})	Fit PT
κ	PT log-kill effect	day^{-1}	4.46×10^{-3} (20.6)	0.0439 (1803)	Fit PT
t_{res}	Half-life of treatment response	day	149 (1.05×10^{-3})	35.4 (0.998)	Fit PT
α_0	Proliferation rate at one cell (BM)	day^{-1}	0.0284	0.0858	Computed
β	Exponential decrease of growth rate (BM)	day^{-1}	1.03×10^{-3}	3.10×10^{-3}	Computed
τ	Dormancy duration	day	133 (4.23)	171 (33.3)	Fit BM
DT_{BM}	BM doubling time at 3 cm	day	173	57	Predicted
μ	Cellular metastatic potential	cell$^{-1}\times$ day^{-1}	2.00×10^{-12} (2.83)	1.02×10^{-12} (71.9)	Fit BM
γ	Fractal (Hausdorff) dimension of the vasculature ($\times\frac{1}{3}$)	-	1	1	Fit BM*
d	Dissemination rate at 3 cm	day^{-1}	0.0282	0.0144	Predicted

Values of the parameters from either the data (Data), the model assumptions (Fixed), estimated from direct computations from diagnosis data (Computed), fit of the TGI model to the PT data (Fit PT) or fit of the metastatic dormancy model to the brain metastases data (Fit BM), or computed from these estimations (Predicted). Values in parenthesis for fitted parameters correspond to the relative standard errors expressed in percent, i.e. $\frac{\text{se}}{\text{est}} \times 100$ where se is the standard error (square root of the covariance diagonal entry, see Materials an methods) and est is the parameter estimate. * = for value of the parameter γ, due to identifiability issues, only five possible values were considered during the fit procedure (see Methods).

Quantitative mathematical modeling of clinical brain metastasis dynamics in non-small cell lung cancer

M. Bilous, C. Serdjebi, A. Boyer, P. Tomasini, C. Pouypoudat, D. Barbolosi, F. Barlesi, F. Chomy and S. Benzekry

Supplementary Material

Table S1: Growth kinetics of primary lung tumors

Figure S1: Comparison of exponential and Gompertz predictions of pre-clinical phase of growth

Figure S2: Gompertz growth predictions of individual brain metastases

Figure S3: Fit of the basic model

Figure S4: Fit of the model with secondary dissemination

Figure S5: Fit of the delay model

Figure S6: Fit of the model with different primary and secondary growth parameters

Figure S7: Fit of the dormancy model for patient 2

Figure S8: Predictions for patient 2

Figure S9: Shape of the objective as a function of μ and γ

Table S1: Growth kinetics of primary lung tumors

Ref.	**All**	**Adenocarcinoma**	**Squamous cell carcinoma**	**Undiff.**
[Jennings et al., 2006]	161 (149)	166 (51)	132 (48)	
[Usuda et al., 1994]	164 $\pm$ 178 (165)	223 $\pm$ 209 (86)	105 $\pm$ 106 (67)	79 $\pm$ 52 (12)
[Arai et al., 1994]	166 (237)	222 (133)	115 (69)	68 (9)
[Mizuno et al., 1984]	136 (50)	178 $\pm$ 157 (23)	103 $\pm$ 78 (22)	111 $\pm$ 57 (5)
[Geddes, 1979]	102 (228)	161 (60)	88 (111)	86 (44)
[Weiss, 1974]	183 $\pm$ 85 (28)	214 $\pm$ 73 (11)	78 $\pm$ 29 (8)	109 $\pm$ 72 (7)
[Spratt and Spratt, 1964]	88 (34)	118 (8)	70 (13)	93 (13)
[Spratt et al., 1963]	112 (22)	269 (7)	93 (6)	90 (9)
[Garland et al., 1963]	162 (41)	222 (7)	126 (22)	123 (9)
[Schwartz, 1961]	78 $\pm$ 48 (13)	72 $\pm$ 77 (2)	79 $\pm$ 46 (11)	
Average*	**143**	**201**	**104**	**91**

Reports of volume doubling times from routinely detected primary NSCLC tumors from the literature (extended from [Detterbeck and Gibson, 2008] and [Friberg and Mattson, 1997]). Note: squamous cell carcinoma = epidermoid carcinoma and undifferentiated carcinoma cell carcinoma = large cell carcinoma. All denotes all histology from lung cancer together (including both NSCLC and small cell carcinomas).
Undiff. = undifferentiated carcinoma.
Mean $\pm$ standard deviation.
In parenthesis are the number of patients.
*Average was computed as the weighted mean of each study with weights the number of patients.

Figure S1: Comparison of exponential and Gompertz predictions of pre-clinical phase of growth

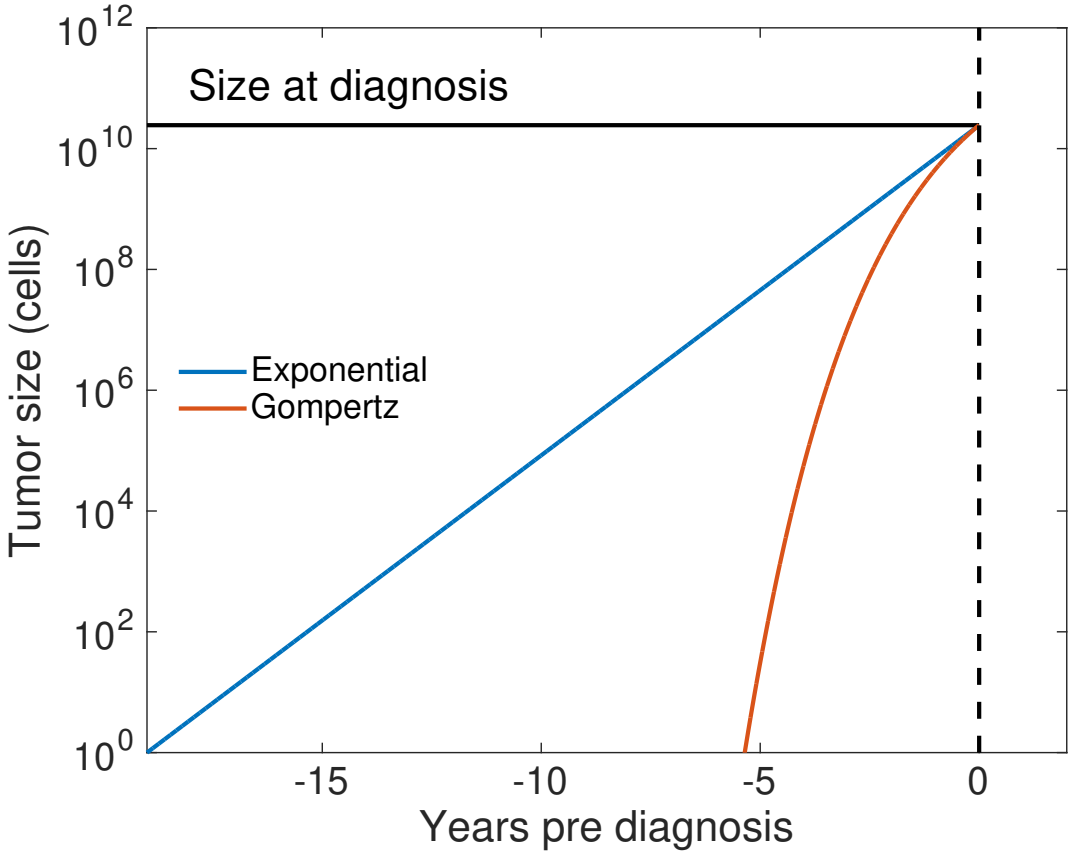
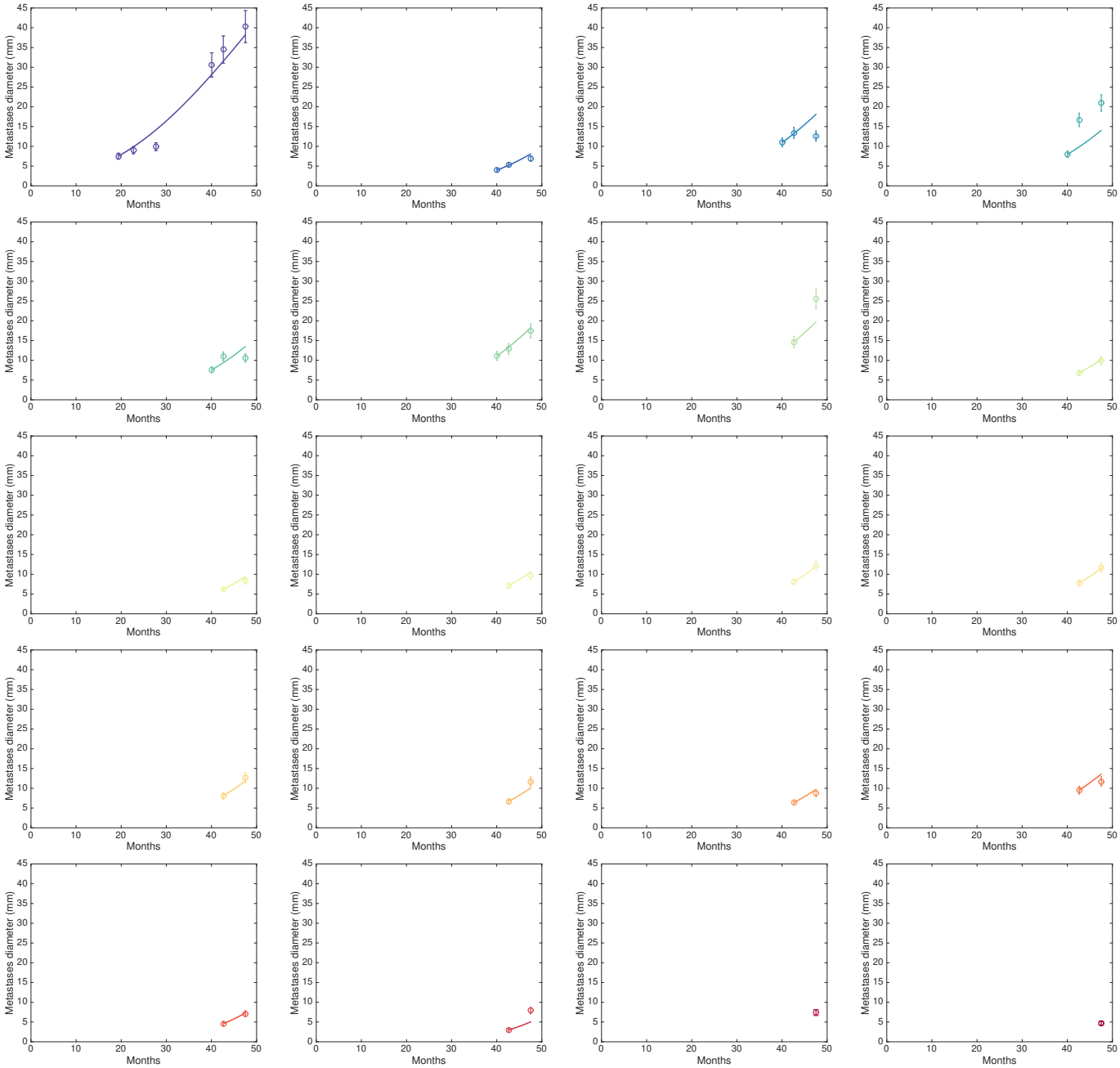
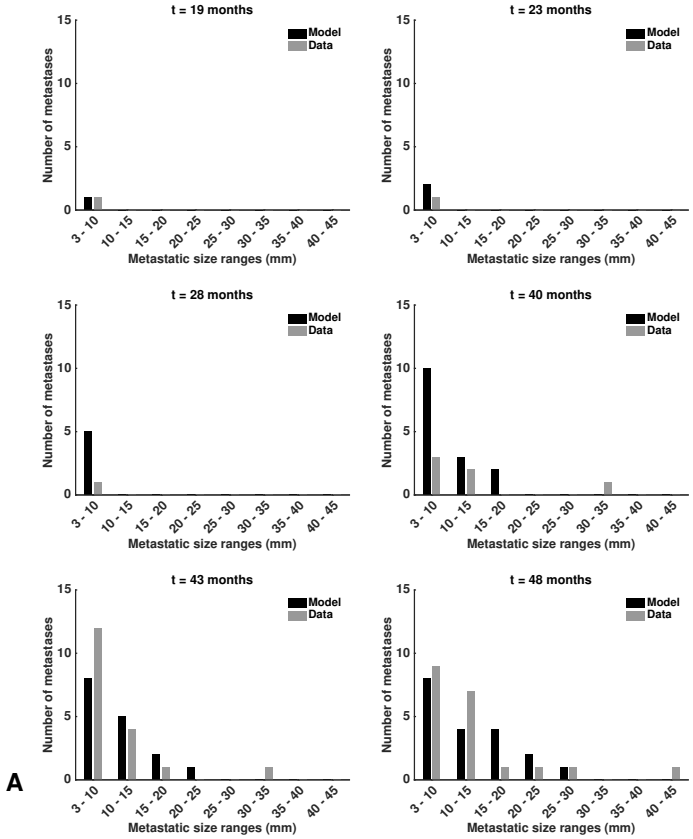


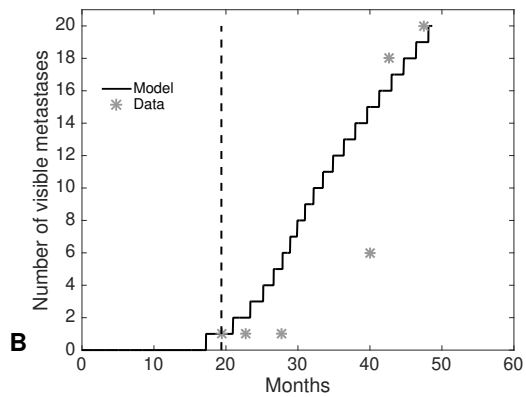
Figure S2: Gompertz growth predictions of individual brain metastases



Comparison between the Gompertz growth model predictions and the lesion size data of individual BMs. The Gompertz growth parameters were determined only from the primary tumor size at diagnosis and histology (adenocarcinoma, see Methods). Apart from the first measurements used as initial conditions, no other future measurements were used to make the predictions. Error bars represent 10% error on the size measurement.

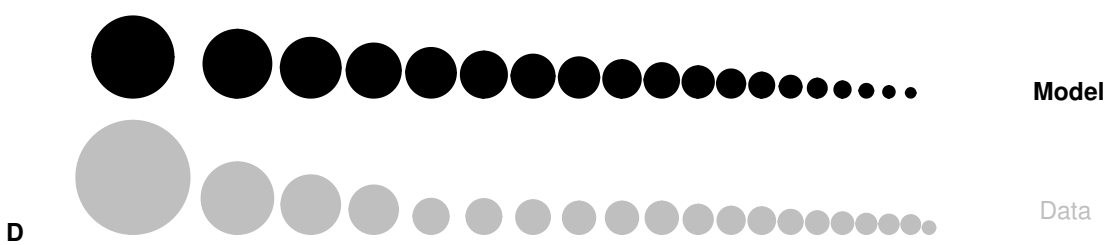
Figure S3: Fit of the basic model





Parameter	Estimate	Unit	Method	Std. error
$\alpha_{0,p}$	0.0284	day^{-1}	Fixed	-
β_p	0.00103	day^{-1}	Fixed	-
α_0	0.0284	day^{-1}	Fixed	-
β	0.00103	day^{-1}	Fixed	-
μ	1.43×10^{-12}	cell$^{-\gamma}\times$ day^{-1}	Fit	3.33
γ	1	-	Variable	-
τ	0	day	Fixed	-
t_d	0	day	Fixed	-

C



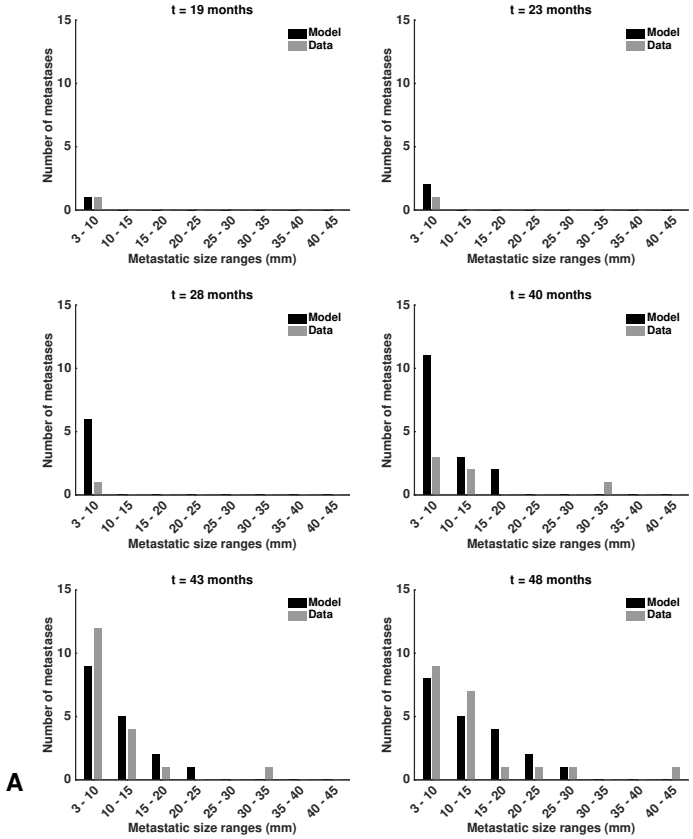
A. Time course of the visible brain metastases (BMs) size distributions during follow-up. Comparison between model calibration and data.

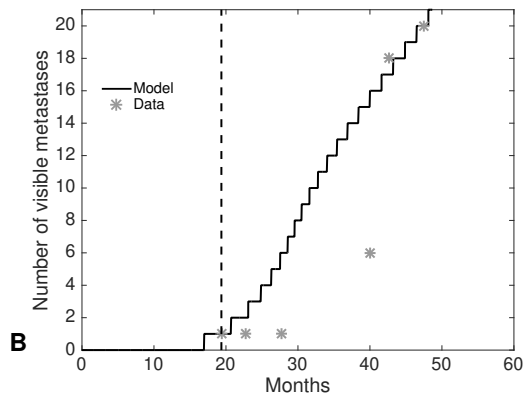
B. Time course of the number of visible BMs.

C. Parameter estimates. Std error = Standard errors expressed in percent.

D. Comparison of the BM size distribution between the model fit and the data at the last time point.

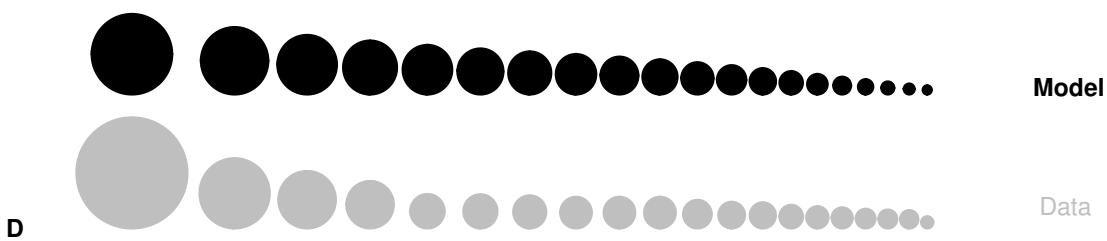
Figure S4: Fit of the model with secondary dissemination





C

Parameter	Estimate	Unit	Method	Std. error
$\alpha_{0,p}$	0.0284	day^{-1}	Fixed	-
β_p	0.00103	day^{-1}	Fixed	-
α_0	0.0284	day^{-1}	Fixed	-
β	0.00103	day^{-1}	Fixed	-
μ	1.5×10^{-12}	cell$^{-\gamma}\times$ day^{-1}	Fit	-
γ	1	-	Variable	-
τ	0	day	Fixed	-
t_d	0	day	Fixed	-



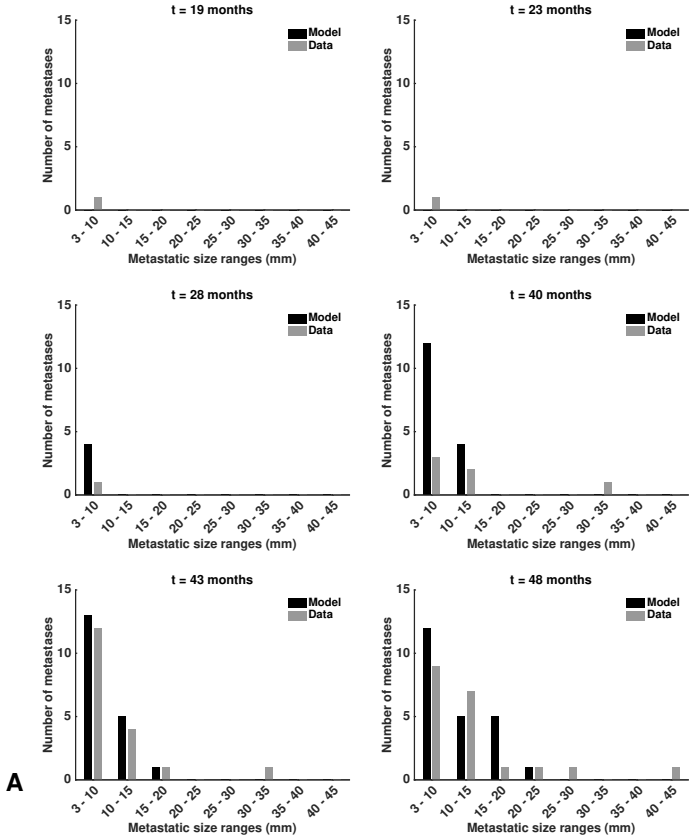
A. Time course of the visible brain metastases (BMs) size distributions during follow-up. Comparison between model calibration and data.

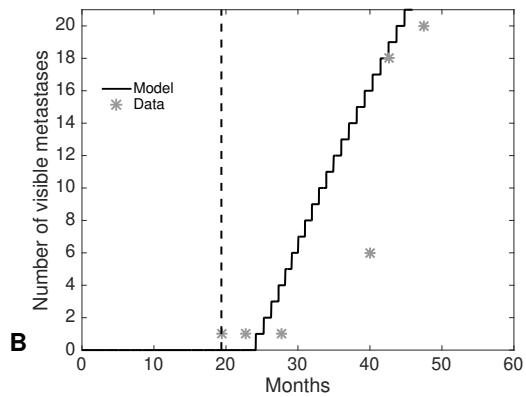
B. Time course of the number of visible BMs.

C. Parameter estimates. Std error = Standard errors expressed in percent.

D. Comparison of the BM size distribution between the model fit and the data at the last time point.

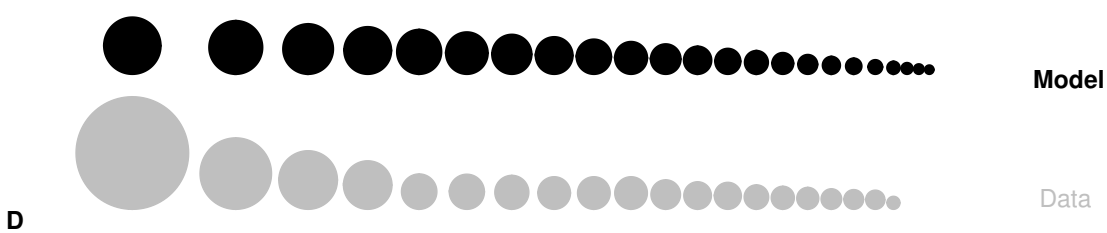
Figure S5: Fit of the delay model





Parameter	Estimate	Unit	Method	Std. error
$\alpha_{0,p}$	0.0284	day^{-1}	Fixed	-
β_p	0.00103	day^{-1}	Fixed	-
α_0	0.0284	day^{-1}	Fixed	-
β	0.00103	day^{-1}	Fixed	-
μ	2.65×10^{-6}	cell$^{-\gamma}\times$ day^{-1}	Fit	10.5
γ	0.4	-	Variable	-
τ	0	day	Fixed	-
t_d	1.76×10^{3}	day	Fit	2.31

C



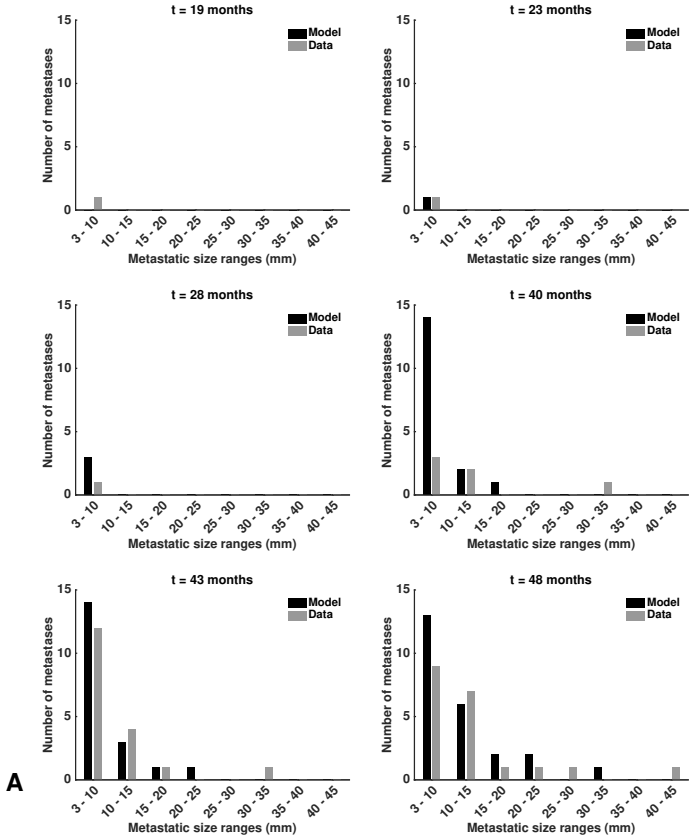
A. Time course of the visible brain metastases (BMs) size distributions during follow-up. Comparison between model calibration and data.

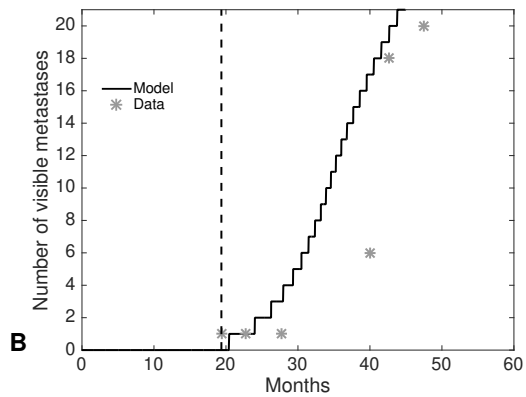
B. Time course of the number of visible BMs.

C. Parameter estimates. Std error = Standard errors expressed in percent.

D. Comparison of the BM size distribution between the model fit and the data at the last time point.

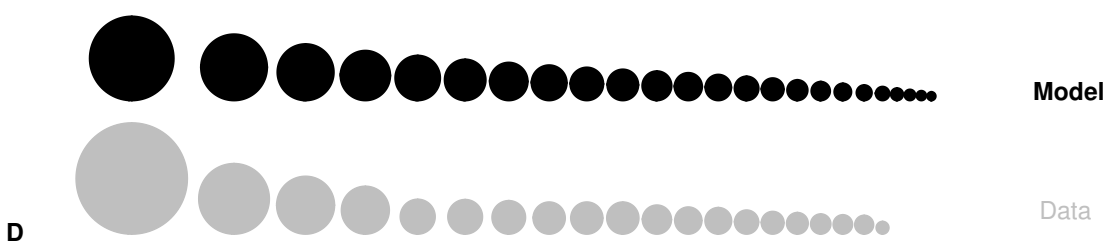
Figure S6: Fit of the model with different primary and secondary growth parameters





C

Parameter	Estimate	Unit	Method	Std. error
$\alpha_{0,p}$	0.0284	day^{-1}	Fixed	-
β_p	0.00103	day^{-1}	Fixed	-
α_0	0.022	day^{-1}	Fit	1.05
β	0.000666	day^{-1}	Fit	0.956
μ	2.07×10^{-12}	cell$^{-\gamma}\times$ day^{-1}	Fit	6.71
γ	1	-	Variable	-
τ	0	day	Fixed	-
t_d	0	day	Fixed	-



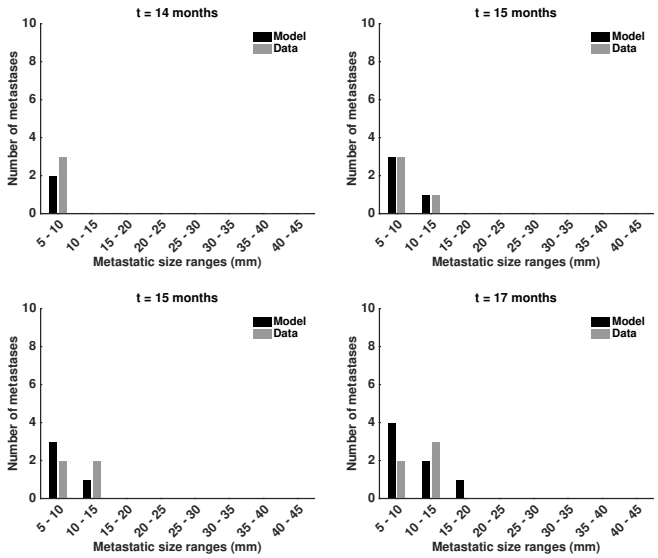
A. Time course of the visible brain metastases (BMs) size distributions during follow-up. Comparison between model calibration and data.

B. Time course of the number of visible BMs.

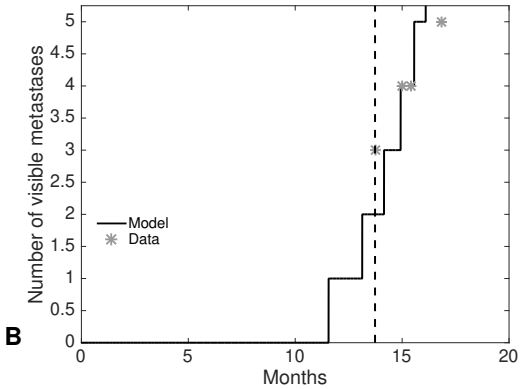
C. Parameter estimates. Std error = Standard errors expressed in percent.

D. Comparison of the BM size distribution between the model fit and the data at the last time point.

Figure S7: Fit of the dormancy model for patient 2



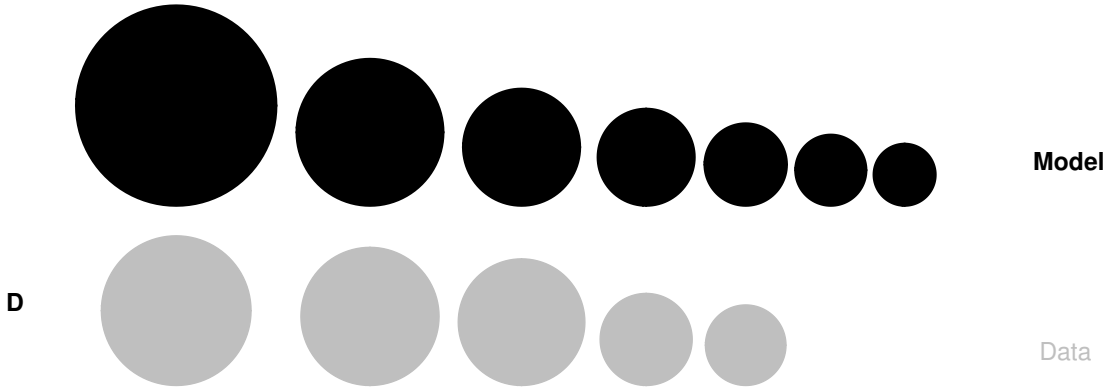
A



B

Parameter	Estimate	Unit	Method	Std. error
$\alpha_{0,p}$	0.0858	day^{-1}	Fixed	-
β_p	0.0031	day^{-1}	Fixed	-
α_0	0.0858	day^{-1}	Fixed	-
β	0.0031	day^{-1}	Fixed	-
μ	1.02×10^{-12}	cell$^{-\gamma}\times$ day^{-1}	Fit	71.9
γ	1	-	Variable	-
τ	171	day	Fit	33.3
t_d	0	day	Fixed	-

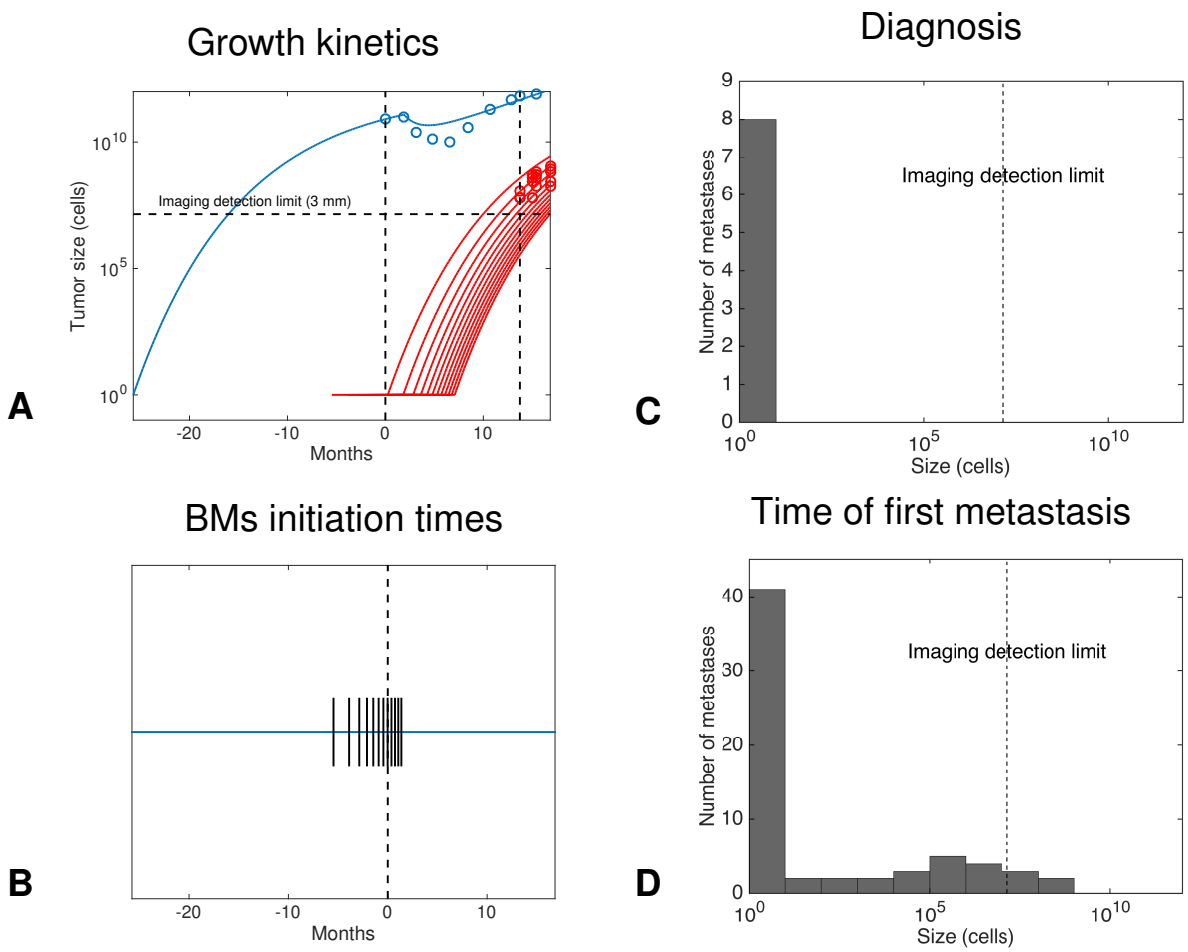
C



D

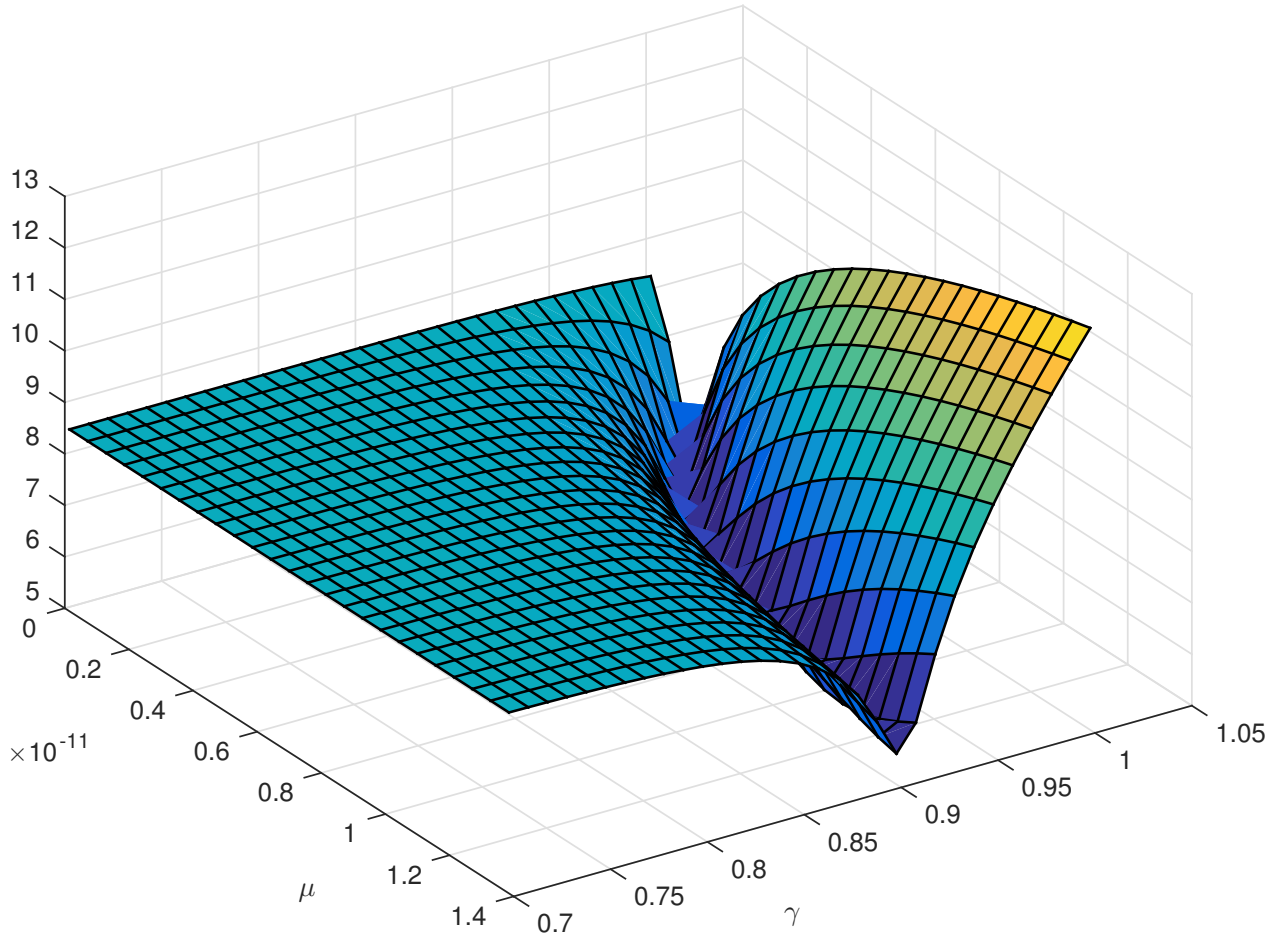
A. Time course of the visible brain metastases (BMs) size distributions during follow-up. Comparison between model calibration and data.
B. Time course of the number of visible BMs.
C. Parameter estimates. Std error = Standard errors expressed in percent.
D. Comparison of the BM size distribution between the model fit and the data at the last time point.

Figure S8: Predictions for patient 2



A. Inferred growth kinetics of the primary tumor (in blue) and the brain metastases (in red). Only brain metastases that will become visible are shown.
B. Model predictions of the initiation times of the brain metastases.
C. Predicted size distribution of the brain metastases at diagnosis.
D. Predicted size distribution of the brain metastases at the time of clinical occurrence of the first one.

Figure S9: Shape of the objective as a function of μ and γ



References

[Arai et al., 1994] Arai, T., Kuroishi, T., Saito, Y., Kurita, Y., Naruke, T., and Kaneko, M. (1994). Tumor doubling time and prognosis in lung cancer patients: evaluation from chest films and clinical follow-up study. Japanese Lung Cancer Screening Research Group. *Jpn J Clin Oncol*, 24(4):199–204.

[Detterbeck and Gibson, 2008] Detterbeck, F. C. and Gibson, C. J. (2008). Turning gray: the natural history of lung cancer over time. *J Thorac Oncol*, 3(7):781–792.

[Friberg and Mattson, 1997] Friberg, S. and Mattson, S. (1997). On the growth rates of human malignant tumors: implications for medical decision making. *J Surg Oncol*, 65(4):284–297.

[Garland et al., 1963] Garland, L. H., Coulson, W., and Wollin, E. (1963). The rate of growth and apparent duration of untreated primary bronchial carcinoma. *Cancer*, 16:694–707.

[Geddes, 1979] Geddes, D. M. (1979). The natural history of lung cancer: a review based on rates of tumour growth. *Br J Dis Chest*, 73(1):1–17.

[Jennings et al., 2006] Jennings, S. G., Winer-Muram, H. T., Tann, M., Ying, J., and Dowdeswell, I. (2006). Distribution of stage I lung cancer growth rates determined with serial volumetric CT measurements. *Radiology*, 241(2):554–563.

[Mizuno et al., 1984] Mizuno, T., Masaoka, A., Ichimura, H., Shibata, K., Tanaka, H., and Niwa, H. (1984). Comparison of actual survivorship after treatment with survivorship predicted by actual tumor-volume doubling time from tumor diameter at first observation. *Cancer*, 53(12):2716–2720.

[Schwartz, 1961] Schwartz, M. (1961). A biomathematical approach to clinical tumor growth. *Cancer*, 14:1272–1294.

[Spratt et al., 1963] Spratt, J. S., SPJUT, H. J., and ROPER, C. L. (1963). The frequency distribution of the rates of growth and the estimated duration of primary pulmonary carcinomas. *Cancer*, 16:687–693.

[Spratt and Spratt, 1964] Spratt, J. S. and Spratt, T. L. (1964). Rates of growth of pulmonray metastases and host survival. *Ann Surg*, 159(2):161–171.

[Usuda et al., 1994] Usuda, K., Saito, Y., Sagawa, M., Sato, M., Kanma, K., Takahashi, S., Endo, C., Chen, Y., Sakurada, A., and Fujimura, S. (1994). Tumor doubling time and prognostic assessment of patients with primary lung cancer. *Cancer*, 74(8):2239–2244.

[Weiss, 1974] Weiss, W. (1974). Tumor doubling time and survival of men with bronchogenic carcinoma. *Chest*, 65(1):3–8.