Computational modeling reveals dynamics of brain metastasis in non-small cell lung cancer and provides a tool for personalized therapy

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*

The authors declare no potential conflicts of interest

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

Abstract

Brain metastases (BMs) are the largest disabling site for non-small cell lung cancers, but are only visible when sizeable. Individualized prediction of the BM risk and extent is a major challenge for therapeutic decision. This study assesses mechanistic models of BM apparition and growth against clinical imaging data.

We implemented a quantitative computational method to confront biologically-informed mathematical models to clinical data of BMs. Primary tumor growth parameters were estimated from size at diagnosis and histology. Metastatic dissemination and growth parameters were fitted to either population data of BM probability (n=183 patients) or longitudinal measurements of number and size of visible BMs (63 size measurements in two patients). Pre-clinical phases from first cancer cell to detection were estimated to 2.1-5.3 years. A model featuring dormancy was best able to describe the longitudinal data, as well as BM probability as a function of primary tumor size at diagnosis. It predicted first appearance of BMs at 14-19 months pre-diagnosis. Model-informed predictions of invisible cerebral disease burden could be used to inform therapeutic intervention.

Author summary: Management of brain metastasis in lung cancer is a major clinical challenge. This study reports on a quantitative modeling analysis of patient-specific time courses of the apparition and growth of brain metastases in lung cancer. Several biological theories were tested by confronting mechanistic mathematical models to clinical data. The best of these models, which features periods of stable metastatic size (dormancy), provides a valuable computational tool for personalized therapy, by informing on the extent of invisible cerebral metastases at various time points during therapeutic management. This information is critical for clinical interventions such as the use of whole brain radiation therapy.

Introduction

Lung cancer is the first 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 a poor prognosis[5]. In addition, BMs are responsible for disabling symptoms decreasing patients' quality of life. Lung cancer is known to be one of the most deleterious in terms of BMs[3,6–9].

The purpose of the current study is to address the following biological and clinical problems using quantitative mathematical modeling: 1) What was the pre-diagnosis course of a patient presenting with NSCLC? When was the first cancer cell initiated? 2) For a patient developing clinically overt BMs, when did cerebral invasion occur? Was it at an early or late stage of the disease (linear versus parallel model of metastasis[10,11])? 3) Were most of the BMs spread by the primary tumor (PT) or was metastasis-to-metastasis spread of significant importance[12–14]? 4) How did the BMs grow in comparison with the PT? At a same or distinct rate? Were there dormancy periods[15,16]? 5) For patients with no BM at diagnosis, what is the risk and extent (if any) of the occult disseminated disease in the brain? When will the BMs appear (if present)? 6) For patients with a limited number of BMs (typically, one to three), what therapeutic strategy to follow, in particular regarding the use of whole-brain radiotherapy (WBRT)?

The latter is particularly relevant since, as of today, the utility of WBRT in the management of BMs from NSCLC is still controversial, due to important neuro-cognitive toxicities[6,17,18]. Several phase III trials were conducted but no firm conclusion applicable for the entire patient population could be drawn[19,20], in particular regarding to the epidermal growth factor receptor (EGFR) mutation status[21,22]. This points to the need of rational tools to decide therapeutic action in a patient-specific way. Similarly, the clinical follow-up and planning of cerebral magnetic resonance imaging (MRI) could highly benefit from individualized predictions of the probability of relapse.

To face these issues, quantitative mathematical modeling may be of considerable help, by providing new insights as well as useful numerical tools in the era of personalized medicine[23,24]. However, despite numerous studies, the majority of the efforts have remained focused on mathematical models at the scale of cell populations[25], and relatively few studies have focused on the metastatic process. Moreover, the confrontation 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]. On the mechanistic side, in the 1970's, Liotta et al. were pioneers in the development of a biologically-based, low-parameterized and experimentally-validated model for all the main steps of the metastatic process[30]. Since then, only a relatively 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 points of view[41–43] and confronted to animal data[37–39].

Apart from notable exceptions[31,34,44,45], the use of mathematical modeling to interrogate clinical data remains very limited. 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. However, experimental procedures are tedious and often only provide access to the total metastatic burden, neglecting its size distribution into distinct metastatic tumors. The latter information requires access to imaging modalities only available in rare occasions in animal experiments[38]. Individual-level clinical data of metastatic development are also challenging to obtain because precise number and size of existing lesions are not routinely reported in medical records. Moreover, patients usually receive treatment soon after diagnosis, which hampers access to the natural course of the disease. Apart from the landmark work of Iwata et al.[31], no study has modeled longitudinal data of individual-level number and size of metastases. Here, we focused on BMs from NSCLC considering that: 1) they are of worse prognosis than other metastatic locations[5], 2) they are easily quantifiable using MRI, 3) due to the blood brain barrier, BMs are often hardly reached by systemic treatments, thus calling for predictive tools and 4) integration of multiple metastatic sites would require substantial improvements of the model beyond the scope of this work.

Grounded on biological knowledge about the organism-scale dynamics of metastatic disease[46], we present here a computational method for calibration (from clinical data) and simulation of: 1) the pre-diagnosis PT growth phase, 2) the BM probability as a function of diagnosis PT size and 3) different biological hypotheses of PT and BM dynamics. The resulting mechanistic model is further employed to infer clinically relevant parameters such as the time of BM appearance and number and size of invisible lesions.

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 models. Exponential growth is the simplest model expressing uncontrolled proliferation and is often adapted to describe tumor growth kinetics during limited observation periods[47]. However, it has been shown that on longer timeframes (typically for volumes to increase 100 to 1000-fold), the specific growth rate of tumors decreases[48], which is well captured by the Gompertz model[48–50]. For calibration of the models, we used primarily the data of the PT size and mean doubling time at diagnosis according to the histology of lesions, retrieved from a meta-analysis from the literature comprising a total of 388 adenocarcinomas and 377 squamous cell carcinomas[47,51] (see supplementary Table S1). For calibration of the Gompertz model (two parameters), we additionally assumed a carrying capacity of 10^{12} cells[10]. For an adenocarcinoma with a median diameter of 35 mm^3 at diagnosis[52], we obtained a pre-diagnosis phase of 19 years in the exponential case versus 5.4 years in the Gompertz case (Figure 1). The first figure seems unrealistic in comparison with previous reports estimating the age of lung primary tumors to be 3-4 years old[53], based on a different method using time to recurrence due to Collins[11]. The Gompertz estimate on the other hand is rather consistent with the literature range. Moreover, the resulting estimate of $\alpha_{0,p}$ –

which can be interpreted as the cellular proliferation rate – was realistic, generating an estimated length of the cell cycle $\left(=\frac{\log(2)}{\alpha_{0,p}}\right)$ of 24.4 hours[54]. We therefore concluded that this model was better adapted to describe the pre-diagnosis natural history of the PT.

Population-level probability of brain metastasis occurrence as a function of primary tumor size can be described by a mechanistic computational model of metastasis

PT size is a major predictive factor of BM in NSCLC[52]. We extracted data from the literature about the quantitative relationship between PT size at diagnosis and BM apparition probability[52], focusing on adenocarcinoma (n = 136) and squamous cell carcinoma (n = 47) because these are the two histological types of the patient data used below. Following a method previously employed for breast cancer metastatic relapse[39,55], the dissemination law we considered (a power law of the PT size, see equation (2)) – combined with the estimated preclinical Gompertz growth of the PT – was able to adequately fit the data, in both histological types (Figure 2). Of note, this result was obtained with a minimal number of parameters to describe the inter-patient heterogeneity in metastatic potential. Namely, apart from the PT size, a population distribution on only the parameter μ was sufficient to describe the data. Consequently, this parameter had a very important coefficient of variation (> 6,000%). The median value μ_{pop} gives a quantitative way to measure the reported higher BM

aggressiveness of adenocarcinomas over squamous cell carcinomas[52] and we found a difference of two $\log_{10}$ orders of magnitude.

A benchmark of biological scenarios against individual longitudinal data of metastatic number and size suggests dormancy

Growth kinetics

We further used our modeling framework – designed to simulate the dynamics of BM apparition and growth (see Methods) – to interrogate patient-specific longitudinal data retrieved from clinical imaging during post-diagnosis follow-up. These consisted in 10 and 11 PT measurements and 47 and 16 BM measurements in two patients (one with an adenocarcinoma, the other with a squamous cell carcinoma). We first analyzed the adenocarcinoma patient to develop the model and used the second patient to validate our findings.

The PT first responded to systemic therapy (EGFR tyrosine kinase inhibitors and chemotherapy) before slowly regrowing (Figure 3A). However, a first distant BM was detected 20 months after diagnosis, which kept growing uncontrolled (Figure 3B). Other BMs appeared during follow-up (Figure 3C), reaching a total of 20 BMs at 48 months, date of last examination (Figure 3D-E).

To model the effect of systemic treatment on the PT, we found that a tumor growth inhibition model[56] (equation (1)) was able to adequately fit the data (Figure 3A).

Interestingly, for BM kinetics, we found that when applying the Gompertz model with parameters calibrated from the method explained above (i.e. only exploiting the PT size at diagnosis and its histology), when using initial BM conditions, the predicted growth curves when using initial conditions matched the observed data surprisingly well (Figure 3B). This indicated that: 1) for this patient, the BMs did not respond to PT therapy, at least during the observed phase, 2) it is reasonable to assume that all BMs grow 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 and 4) this growth rate is reasonably well estimated from the method we proposed.

Quantitative assessment of five theories 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 apparition 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). It consists in describing the population of BMs by means of a time-dependent size-structured density. The relevant quantity to be compared with the model is then the BM cumulative size distribution (Figure 3E).

We first asked whether an elementary base model was able to reproduce the data. It consisted mostly in the assumptions of 1) same growth parameters for

the PT and the BMs and 2) no secondary dissemination (i.e. BMs spread only by the PT). The best fit of this model was inaccurate (Figure S1), suggesting that more intricate phenomena were at play. Therefore, we improved the base model into four more intricate scenarios (Figure 4) to be tested against the data: 1) secondary dissemination, i.e. the ability of BMs to spread BMs themselves[31], 2) 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[10]), 3) differential growth, i.e. different growth parameters for the PT and the BMs and 4) dormancy, i.e. the ability of disseminated cells to survive as single cells or as a small size bulk for a given period before resuming growth[15,16,57].

The models exhibited differential descriptive power, as quantified by the best-fit value of the objective function (Table 1). Interestingly, inclusion of secondary dissemination yielded similar results as the base model (Figure S2), suggesting that if this process does occur in the reality, then it does not affect significantly the time course of visible BMs. Indeed, under this model, even at the last time point (48 months post-diagnosis), there were only 12 second-generation BMs representing a small total burden of 18,700 cells with the largest BM comprising 16,400 cells ($\simeq 0.32$ mm).

Adjunction in the model of a delay for BM spreading initiation significantly improved the fit (Figure 5C), with a very large inferred value of the delay t_d (4.8 years after the first cancer cell, i.e. 6 months before diagnosis). However, the

improvement was not as good as for the two remaining models, in particular on the cumulative size distribution (Figure S3A). Consideration of different growth parameters between the PT and the BMs led to 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 S4). Of note, the estimated BM growth parameters remained close to the PT ones. The dormancy model also achieved accurate goodness-of-fit, with an estimated dormancy time of $\tau = 133$ days (Figure 5 and Table 2). Together, these two models were the best to describe both the dynamics of BM apparition (Figure 5A-B) and their size distribution (Figure 5C). 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 to 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 latter for further predictions. It should be noted however that despite good description overall of the time dynamics of cumulative size distributions, the size of the largest metastasis was underestimated (Figure 5C). Clinically relevant inferences about the actual time of appearance of the BMs were similar.

Collectively, our methods allowed a quantitative comparison of several theories of metastatic development, suggesting rejection of the base, secondary and delay models alone while the dormancy and different growth models were able to explain the data.

Clinically relevant simulations of the disease course reveals time of brain metastases onset

From the quantitative calibration of the dormancy model to the data, several inferences 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 can be the sign 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 ($\mu = 2 \times 10^{-12}$ versus $\mu_{pop} = 1.39 \times 10^{-11}$, see Table 2), giving further support to our population approach. All parameters were estimated with excellent precision (standard errors < 5%, Table 2).

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 present but 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 could be conducted. 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 6A-B). In fact, at the time of diagnosis, our model predicted the presence of 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 6C. Notably, when the PT was at the size of the visibility threshold, no metastasis was predicted to have occurred yet. This suggests that if the disease had been detected through systematic screening, BM occurrence might have been prevented provided the tumor had been operable at this time. Of interest to radiotherapists, the amount of predicted BMs present at time of appearance of the first BM was already of 28 tumors, the largest one being 2.79 mm large and with an overall BM burden of 1.7×10^7 cells. Therefore, provided that neuro-cognitive risks would be acceptable, the model would recommend pan-cerebral intervention rather than localized intervention only.

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

Validation in 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

growth rate (α_1) as compared to the first patient. The qualitative structure of the model confirmed its descriptive power by being able to give an accurate description of both the treated PT and BMs dynamics (Figure S5-6), while the "base" and "secondary dissemination" models were not. Several parameters appeared to be patient-specific, such as the dormancy duration τ, estimated to 171 days. Due to the lower number of data points available for this patient, parameters identifiability was found worse, but still acceptable for the dissemination and BM growth model (Table 2).

Resulting clinical predictions were distinct (Figure S6), emphasizing the patient-specific nature of BM dynamics. The first BM occurred 14 months after PT onset, 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 BMs being still dormant at the time of diagnosis. Whereas the first BM appeared sooner ion the second patient, our model was able to describe the lower mass of cells thanks to the dormancy parameter.

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 allows: 1) to infer the disease age from the PT size and histology, 2) to test multiple scenarios of metastatic dissemination and colonization against macroscopic data available in the clinic – which suggested the presence of prolonged periods of dormancy–, and 3) to infer times of BM initiations and number and size of invisible lesions.

Estimation of the duration of the preclinical PT growth has important implications in terms of prediction of BMs, since BM is 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 estimating similar unrealistically long pre-diagnosis periods under exponential growth (up to 54 years for tumors detected during screening programs[51]). 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[53].

Importantly, we were then able to lever this description of the natural PT growth into a mechanistic model able to describe the probability of BM occurrence. Our model then provided a quantitative theory for the reported differences in BM

between adenocarcinomas and squamous cell carcinomas in terms of a difference in the cell-scale dissemination parameter μ.

More generally, we believe that our computational platform provides a way to translate biological findings into clinically useful numerical tools. However, in order to provide robust inference, the complexity of the model 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. The metastatic potential μ for instance is the product of several cell-scale probabilities relating to the multiple steps of the metastatic cascade[58]. Interestingly, the median value inferred from a population analysis based on probabilities of BM ($\mu_{pop} = 1.39 \times 10^{-11}$ and 1.76×10^{-13}) matched the ones that were found from analysis of patient-specific data ($\mu = 2 \times 10^{-12}$ and 1.02×10^{-12}), given their variability. Higher metastatic ability found in the patient 2 could be due to histology or to the EGFR mutational status, known to impact on the BM aggressiveness[3,22,59,60]. Interestingly, while relying on distinct modeling techniques, the value we obtained is in the biologically realistic range derived by others using stochastic evolutionary modeling[34,44].

It is an open debate in the literature to determine the value of γ in the dissemination law (2). Some use $\gamma = \frac{2}{3}$, arguing that cells able to leave are located at the surface of the PT[31,37]. Others found that small values of γ close to zero were most appropriate, using this finding as support for the cancer stem cells theory (constant amount of metastatic susceptible cells within the PT

regardless of its size)[61]. We have previously demonstrated that the value of γ was not identifiable from longitudinal data of total metastatic burden[39]. Unfortunately, we reached the same conclusion from the data available in the present study, and used $\gamma = 1$ based on a parsimony argument, as done by others who don't consider the spatial repartition of metastatically-able cells[34,44]. Further insights about the spatial repartition of metastatic clones might emerge from bulk sequencing studies of combined PT and metastases[12].

Second-order phenomena that were ignored here for the sake of identifiability but nevertheless could impact on systemic dynamics include tumor-tumor interactions, either through soluble circulating factors[62] or by exchange of tumor cells between established lesions[12,63]. 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[64,65].

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 (other than on the PT or surgery) is not included. We intend to address this in future work, in particular for optimizing and personalizing the design of combination therapies[24,66,67]. Moreover, the current study needs to be extended to a larger number of patients to decipher general trends in BM patterns of NSCLC patients.

In order to translate our findings into a clinically usable tool, further methods need to be developed to calibrate the small number of parameters of the model from data already available at diagnosis or at the first BM occurrence. To this respect, in addition to routine clinico-demographic features, molecular gene expression signatures[68] (for parameter μ, for instance) as well as radiomics features predictive of metastatic relapse[69] (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[52] 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 (International Review Board of the French learned society for respiratory medicine, reference number 2015-041), according to French law. The data from the Bordeaux patient was collected during a retrospective study occurring prior to the 2016 Jardé act requiring formal approval by an ethics

committee. It was nevertheless approved by the internal Bergonie research college. The data were analyzed anonymously.

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 – was assumed to follow the Gompertz growth model[49,50], i.e.

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

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[70]. After treatment start (T_d), the primary tumor size was assumed to follow a tumor growth inhibition model[56] consisting of: 1) exponential growth (rate α_1), 2) log-kill effect of the therapy (efficacy parameter κ)[71] 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

Base model

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[72]. The multiple steps of the metastatic cascade[58] are aggregated into a dissemination rate with expression:

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

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[73]. 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 (base model) or distinct (different growth model) 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:

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

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

$$f(t,s) = N(t - t(s)), \tag{5}$$

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.

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

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

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:

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

Secondary dissemination

In the previous model formulations, 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]:

$$\begin{cases} \partial_t \rho(t,s) + \partial_s\big(\rho(t,s)g(s)\big) = 0 & t \in (0,+\infty),\ \ s \in (S_0,+\infty) \\ g(S_0)\rho(t,S_0) = d\left(S_p(t)\right) + \int_{S_0}^{+\infty} d(s)\rho(t,s)ds & t \in (0,+\infty) \\ \rho(0,s) = 0 & s \in (S_0,+\infty) \end{cases} \tag{6}$$

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].

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 the base model case, the appearance time of the i-th BM T_i is defined by

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

The size of the i-th BM $S_i(t)$ is then defined, for $t > T_i$, by:

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

For links between stochastic (Poisson process) and continuous versions of the Iwata model, the reader is referred to[74].

Models' fits and parameters estimation

Parameters calibration for the primary tumor growth

To parameterize the Gompertz function defining the PT growth, two parameters need 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[10,29] 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[47,51]). 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[52] describing the probability of BM in a population of lung adenocarcinoma patients, we employed a previously described methodology[39,55]. Briefly, we considered that the probability of developing BM after diagnosis was the probability of having already BM at diagnosis, i.e. $\mathbb{P}(N(T_d) > 1)$. We fixed the value of the PT growth parameters as described above from the cohort histology (adenocarcinoma) 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 $S^1, \ldots, S^6$ 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)[75].

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:

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

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 "base 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:

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

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

Parameters identifiability

Standard errors can be computed from this statistics' covariance matrix, given by[76]:

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

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

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

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 base 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. Villà, S. *et al.* Validation of the new Graded Prognostic Assessment scale for brain metastases: a multicenter prospective study. *Radiat Oncol* **6,** 23 (2011).

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

11. Collins, V. P., Loeffler, R. K. & Tivey, H. Observations on growth rates of

human tumors. *Am J Roentgenol Radium Ther Nucl Med* **76,** (1956).

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

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

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

15. 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).

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

17. Owonikoko, T. K. *et al.* Current approaches to the treatment of metastatic brain tumours. *Nat Rev Clin Oncol* **11,** 203–222 (2014).

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

19. 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).

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

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). doi:10.1038/nrclinonc.2015.204

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

45. Hanin, L. & Rose, J. Uncovering the natural history of cancer from post-mortem cross-sectional diameters of hepatic metastases. *Math Med Biol* **33,** 397–416 (2016).

46. Valastyan, S. & Weinberg, R. A. Tumor metastasis: molecular insights and

evolving paradigms. *Cell* **147,** 275–92 (2011).

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

48. 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).

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

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

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

52. 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).

53. 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).

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

55. Barbolosi, D. *et al.* Modélisation du risque d'évolution métastatique chez les patients supposés avoir une maladie localisée. *Oncologie* **13,** 528–533 (2011).

56. 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).

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

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

127, 679–695 (2006).

59. 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).

60. 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).

61. Hanin, L., Seidel, K. & Stoevesandt, D. A "universal" model of metastatic cancer, its parametric forms and their identification: what can be learned from site-specific volumes of metastases. *J Math Biol* **72,** 1633–1662 (2016).

62. 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).

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

64. 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).

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

66. 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).

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

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

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

70. 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).

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

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

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

74. 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

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

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

Figure and table legends

Figure 1: Comparison of exponential and Gompertz predictions of pre-clinical phase of growth

Figure 2: Probability of brain metastasis as a function of the primary tumor size

Fit of our mechanistic model for probability of BM to data from Mujoomdar et al.[52]

A. Adenocarcinoma data (n=136). Inferred values for the distribution of μ: $\mu_{pop} = 1.39 \times 10^{-11}$ cell^{-1} day^{-1} and $\mu_{\sigma} = 1.24 \times 10^{-9}$ cell^{-1} 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} day^{-1} and $\mu_{\sigma} = 1.11 \times 10^{-11}$ cell^{-1} day^{-1}

Figure 3: 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, and initial condition the time of apparition 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).

E. Cumulative size distribution of brain metastases 48 months after diagnosis.

EGFR = Epithelial Growth Factor Receptor, TKI = Tyrosine Kinase Inhibitor, mTKI = maintenance Tyrosine Kinase Inhibitor, CT = (cytotoxic) chemotherapy.

Figure 4: Schematic of the models and investigated hypotheses

Figure 5: Fit of the dormancy model

A. Time course of the visible brain metastases (BMs) cumulative size distributions during follow-up. Black stars = model. Red circles = data.

B. Time course of the number of visible BMs

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

Figure 6: Clinically relevant predictions and inferences

A. Inferred growth kinetics of the primary tumor (in blue) and the brain metastases.

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.

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

Table 2 : Patient-specific parameters

Values of parameters from either the data, the model assumptions, estimated from fitting the model to the data or computed from these estimations. Values in parenthesis only apply to parameters subject to estimation through the minimization procedure and correspond to the relative standard errors expressed in percent, i.e. $\frac{\boldsymbol{se}}{\boldsymbol{est}} \times \mathbf{100}$ where $\boldsymbol{se}$ is the standard error (square root of the covariance diagonal entry, see Materials an methods) and $\boldsymbol{est}$ is the parameter estimate.

Figures and Tables

Figure 1: Comparison of exponential and Gompertz predictions of pre-clinical phase of growth

Figure 2: Probability of brain metastasis as a function of the primary tumor size

Figure 3: Data of primary tumor and brain metastases in a patient with non-small cell lung cancer

Figure 4: Schematic of the models

Figure 5: Benchmark of the models against the data

Figure 6: Clinically relevant predictions and inferences

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

Table 2: Patient-specific parameters estimates

Figure 1: Comparison of exponential and Gompertz predictions of pre-clinical phase of growth

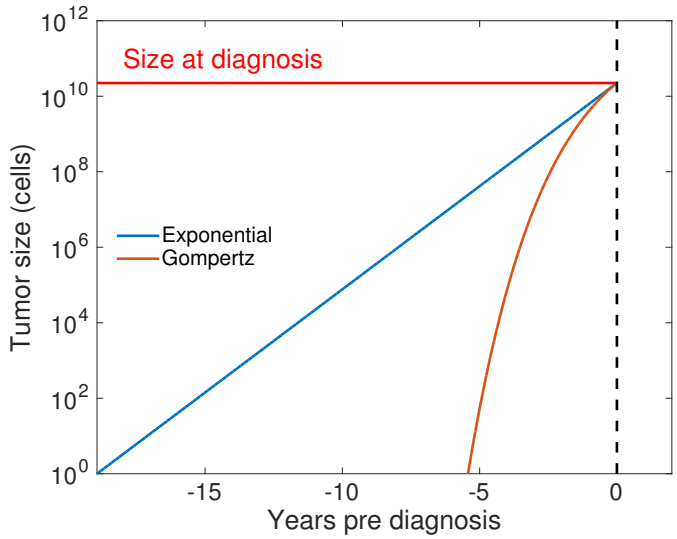


Figure 2: Probability of brain metastasis as a function of the primary tumor size

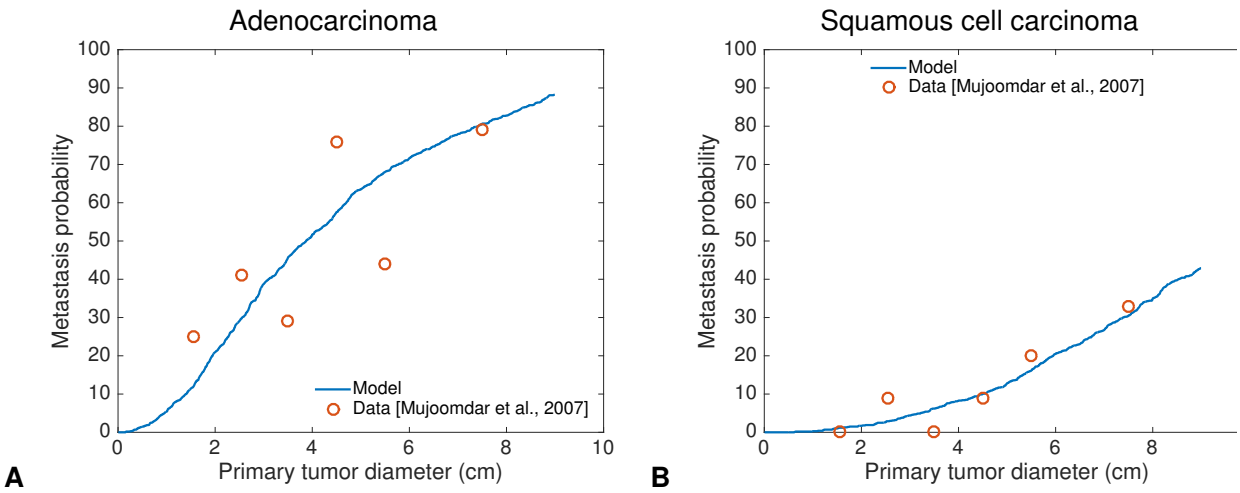


Figure 3: Data of primary tumor and brain metastases in a patient with non-small cell lung cancer

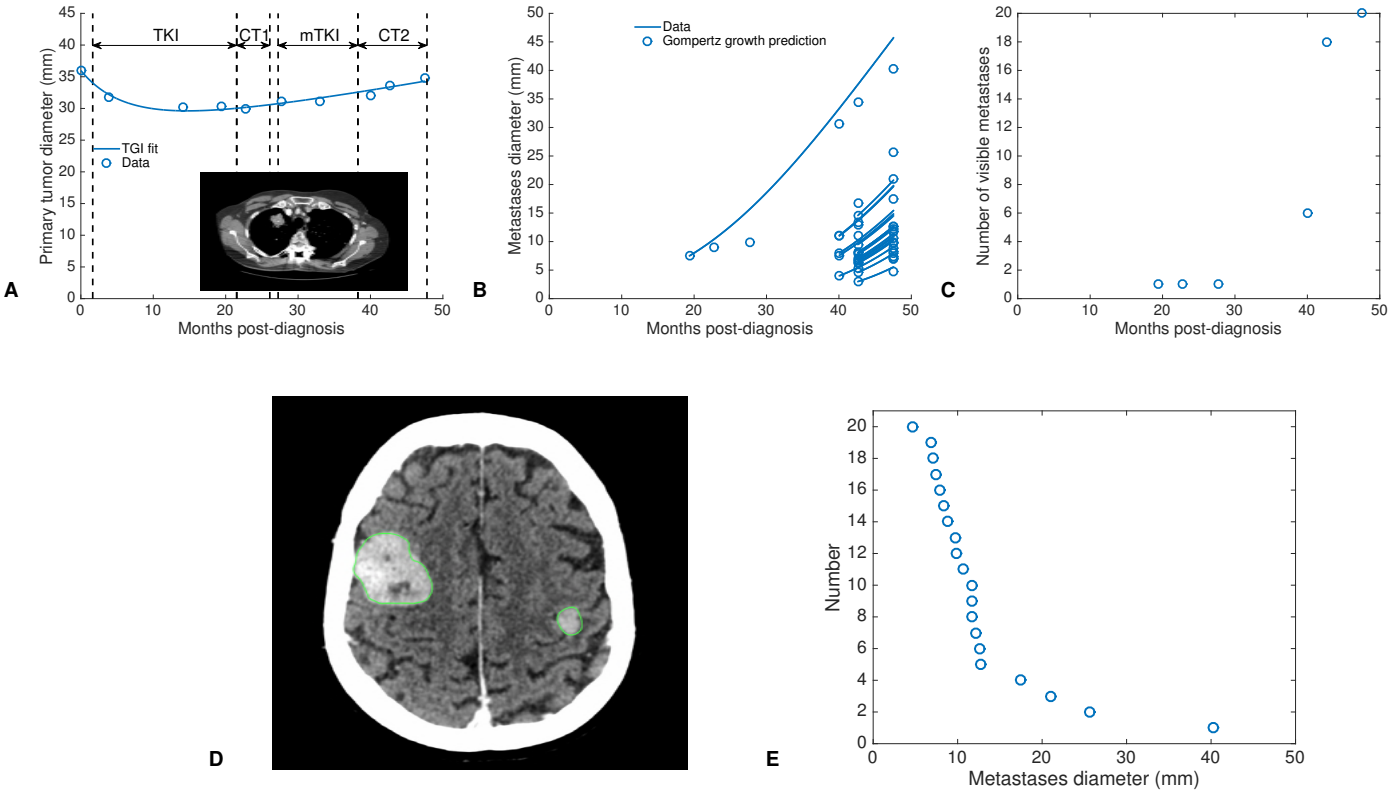


Figure 4: Schematic of the models

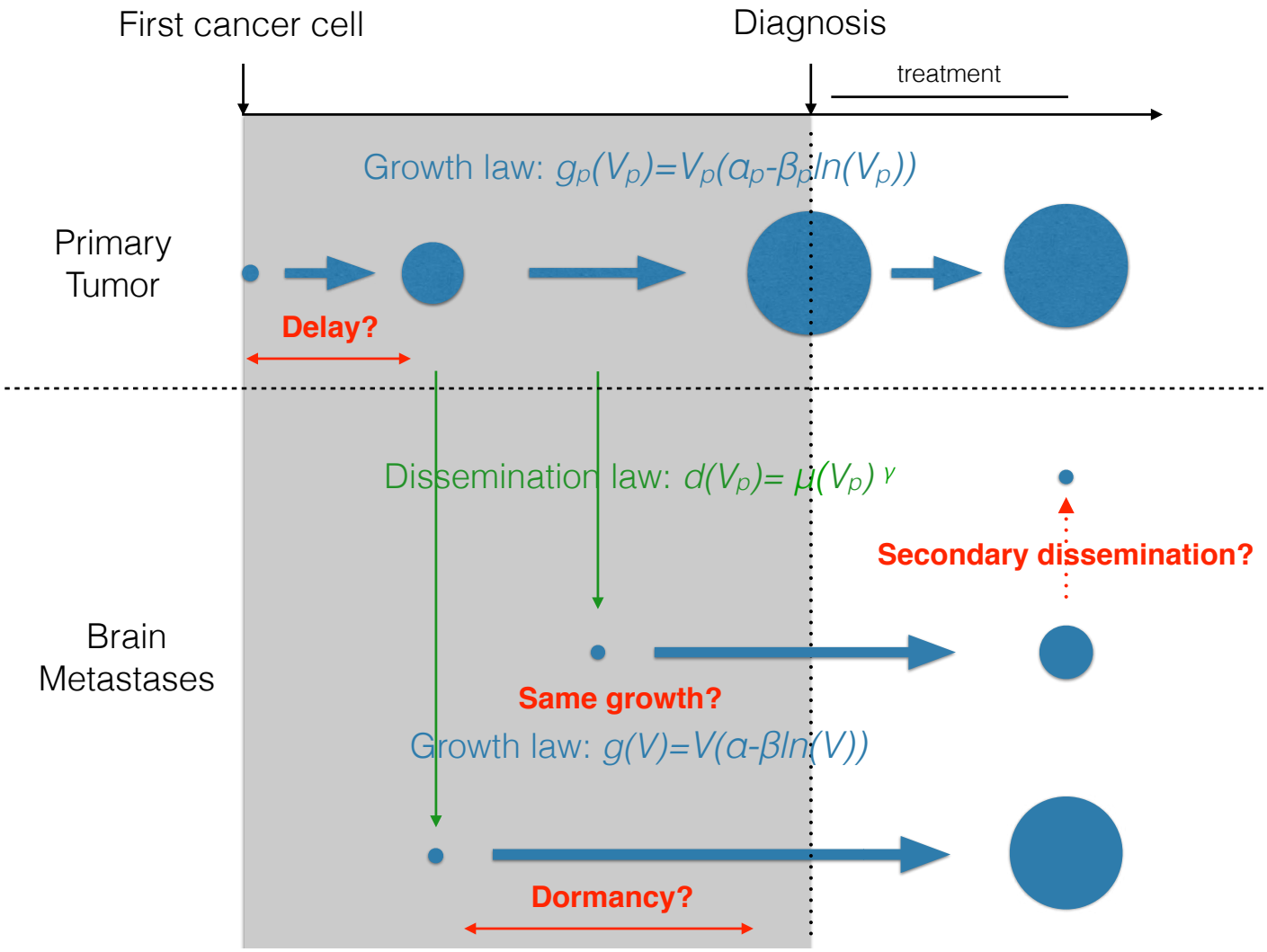


Figure 5: Benchmark of the models against the data

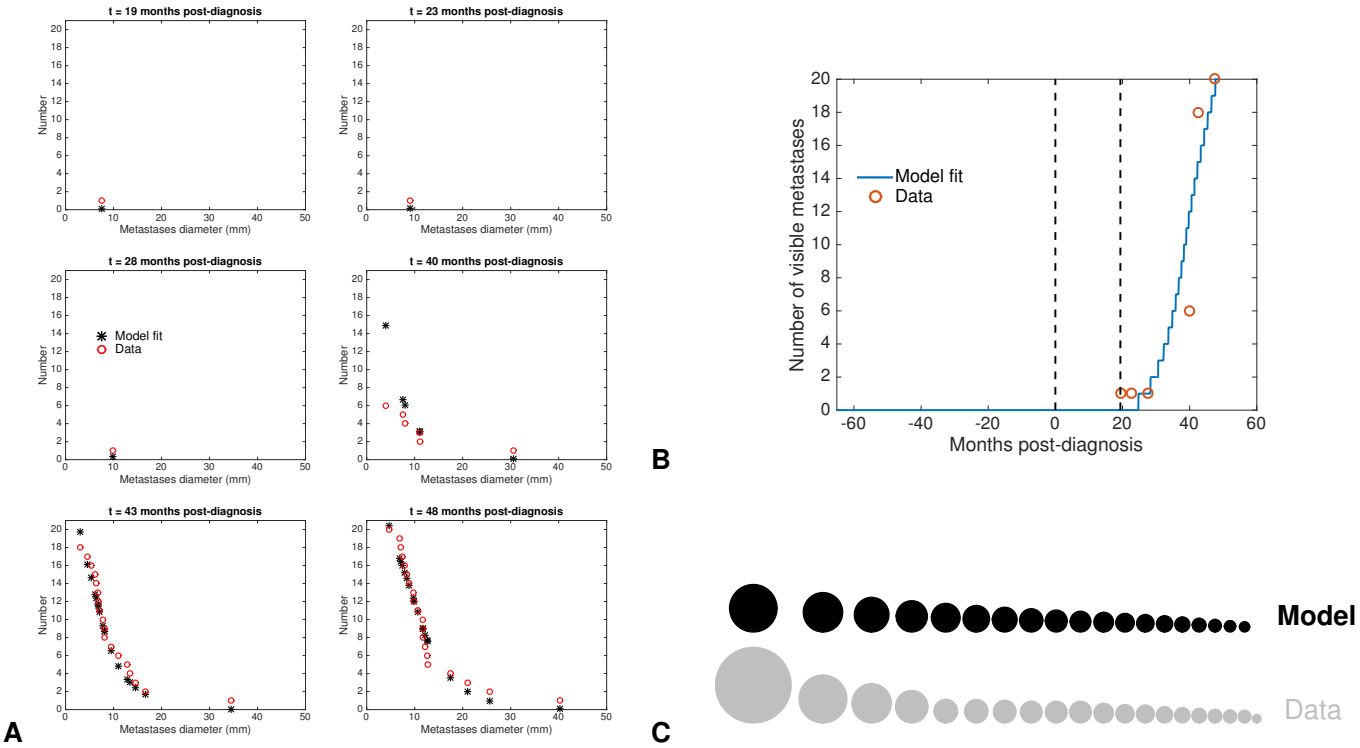


Figure 6: Clinically relevant predictions and inferences

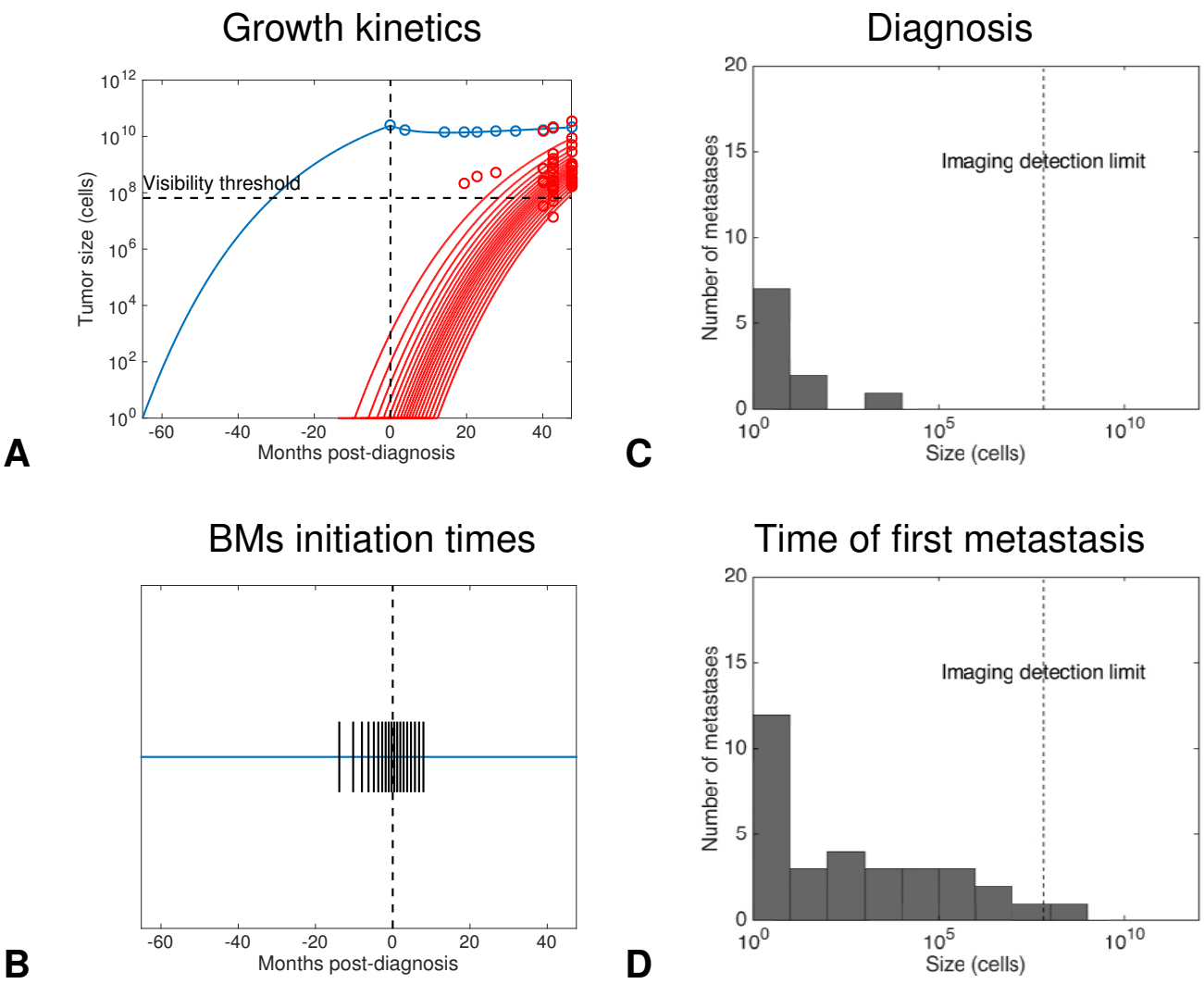


Table 1: Minimal value of the objective function obtained when fitting each of the models to the data

Model	Patient 1	Patient 2
Base	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

Table 2: Patient-specific parameters estimates

Parameter	Meaning	Unit	Patient 1	Patient 2
S_d	PT size at diagnosis	mm	36.0	53.7
Histology			Adenocarcinoma	Squamous cell carcinoma
T_d	Age of the PT at diagnosis	year	5.3	2.1
Visibility threshold		mm	5	5
$S_{0,p}$	Initial size of the PT	cell	1	1
S_0	Initial size of a BM	cell	1	1
$\alpha_{0,p}$	Proliferation rate at one cell (PT)	day^{-1}	0.0284	0.0858
β_p	Exponential decrease of growth rate (PT)	day^{-1}	1.03×10^{-3}	3.10×10^{-3}
DT_{PT}	PT doubling time at 3 cm	day	173	57
α_1	PT growth rate during relapse	day^{-1}	5.72×10^{-4} (23.6)	0.0101 (5.28×10^{-3})
κ	PT log-kill effect	day^{-1}	4.46×10^{-3} (20.6)	0.0439 (1803)
t_{res}	Half-life of treatment response	day	149 (1.05×10^{-3})	35.4 (0.998)
α_0	Proliferation rate at one cell (BM)	day^{-1}	0.0284	0.0858
β	Exponential decrease of growth rate (BM)	day^{-1}	1.03×10^{-3}	3.10×10^{-3}
τ	Dormancy duration	day	133 (4.23)	171 (33.3)
DT_{BM}	BM doubling time at 3 cm	day	173	57
μ	Cellular metastatic potential	cell$^{-1}\times$ day^{-1}	2.00×10^{-12} (2.83)	1.02×10^{-12} (71.9)
γ	Fractal (Hausdorff) dimension of the vasculature ($\times\frac{1}{3}$)	-	1	1
d	Dissemination rate at 3 cm	day^{-1}	0.0282	0.0144

Supplementary Figures

Table S1: Growth kinetics of primary lung tumors

Figure S1: Fit of the baseline model

Figure S2: Fit of the model with secondary dissemination

Figure S3: Fit of the delay model

Figure S4: Fit of the model with different primary and secondary growth parameters

Figure S5: Fit of the dormancy model for patient 2

Figure S6: Clinical inference for patient 2

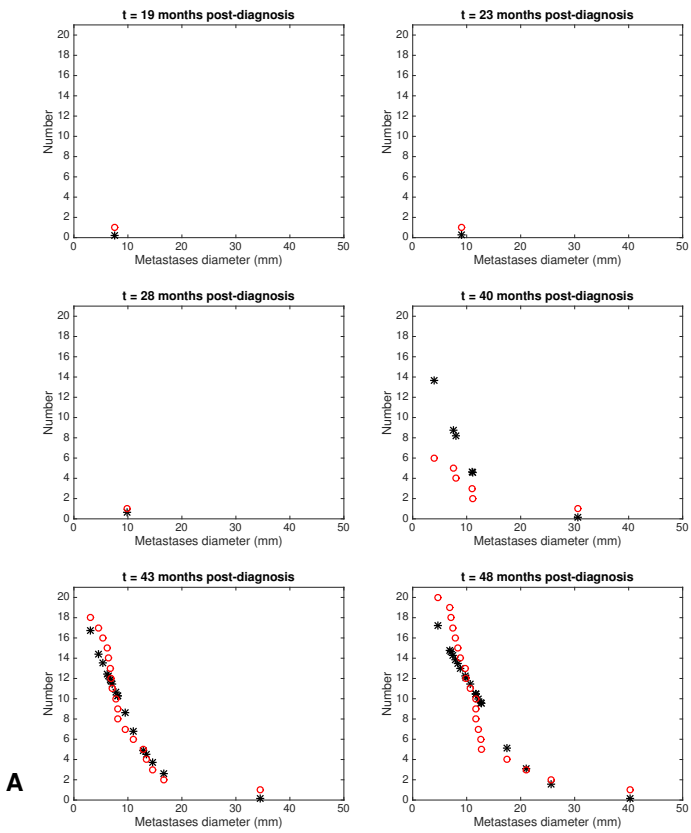
Figure S7: Shape of the objective as a function of μ and γ

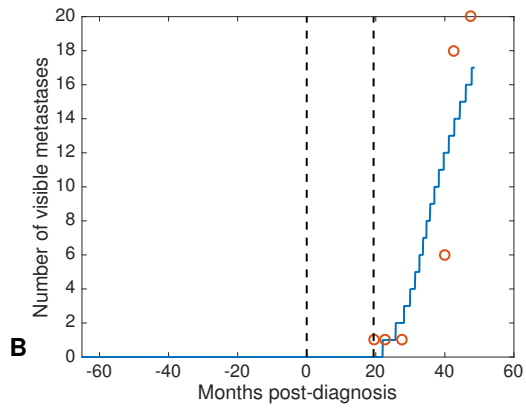
Table S1: Growth kinetics of primary lung tumors

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.

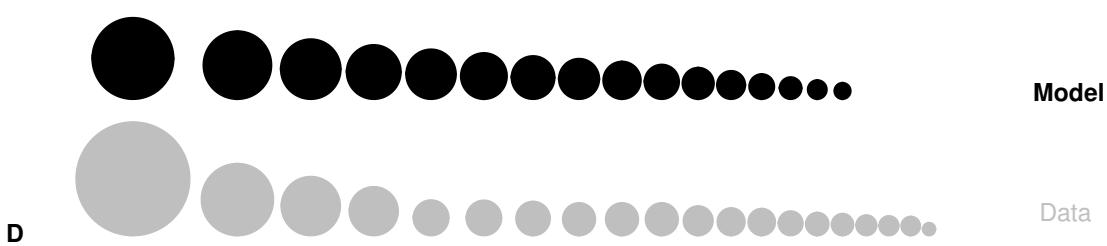
Figure S1: Fit of the baseline model





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	1.43×10^{-12}	cell$^{-\gamma}\times$ day^{-1}	Fit	3.33
γ	1	-	Variable	-
τ	0	day	Fixed	-
t_d	0	day	Fixed	-



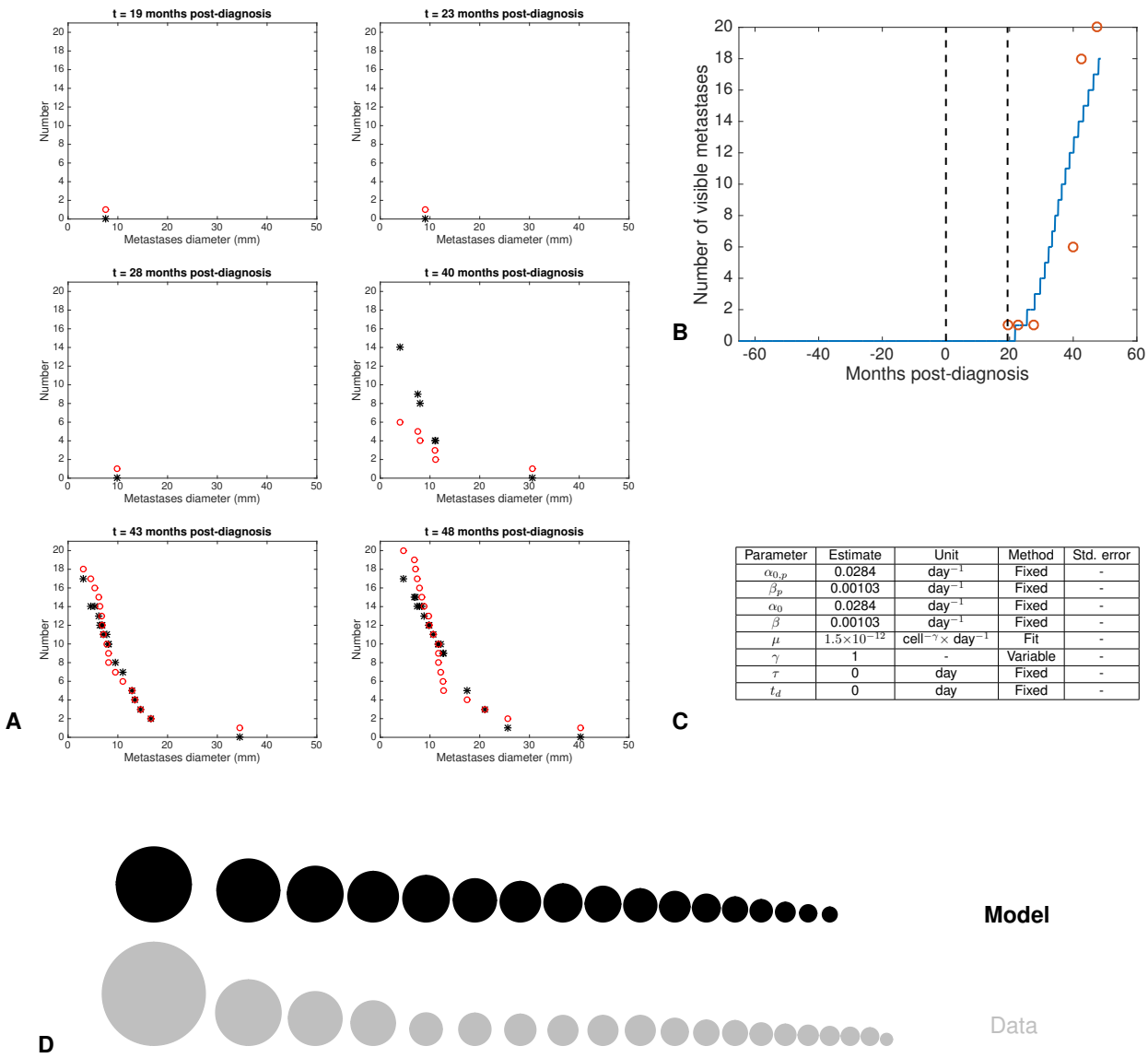
A. Time course of the visible brain metastases (BMs) cumulative size distributions during follow-up. Black stars = model. Red circles = data.

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 (48 months post-diagnosis).

Figure S2: Fit of the model with secondary dissemination



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	-

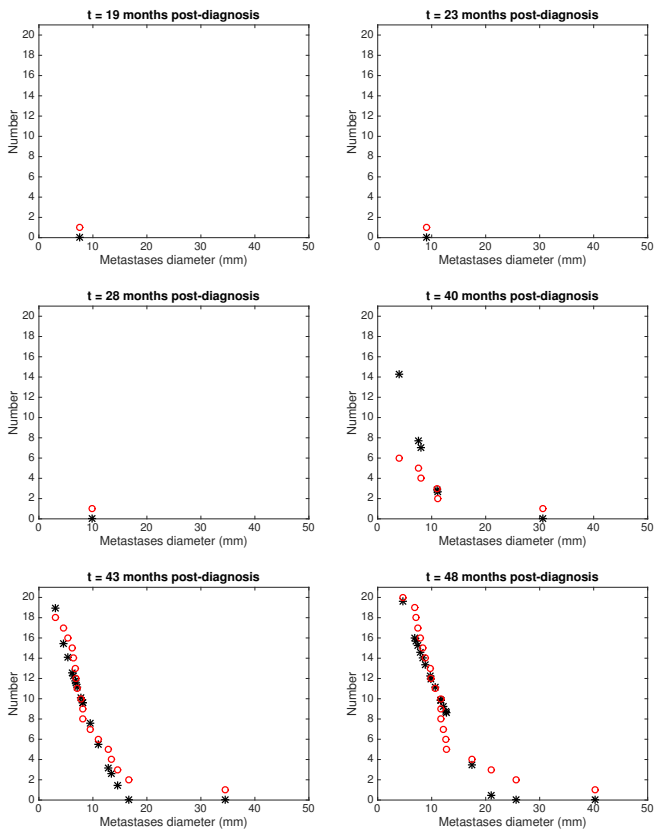
A. Time course of the visible brain metastases (BMs) cumulative size distributions during follow-up. Black stars = model. Red circles = data.

B. Time course of the number of visible BMs.

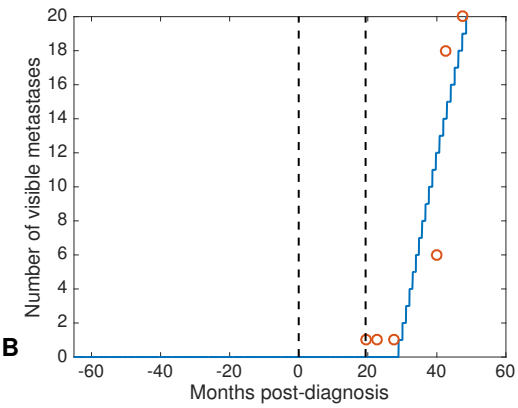
C. Parameter estimates. Std error = Standard errors expressed in percent.

D. Comparison of the BM size distribution between the model fit and the data at the last time point (48 months post-diagnosis).

Figure S3: Fit of the delay model



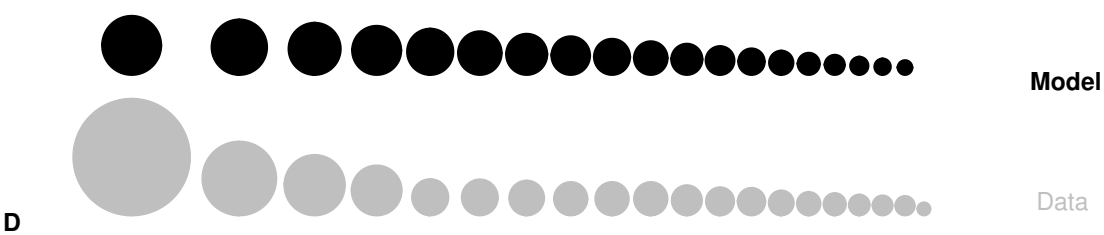
A



B

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



D

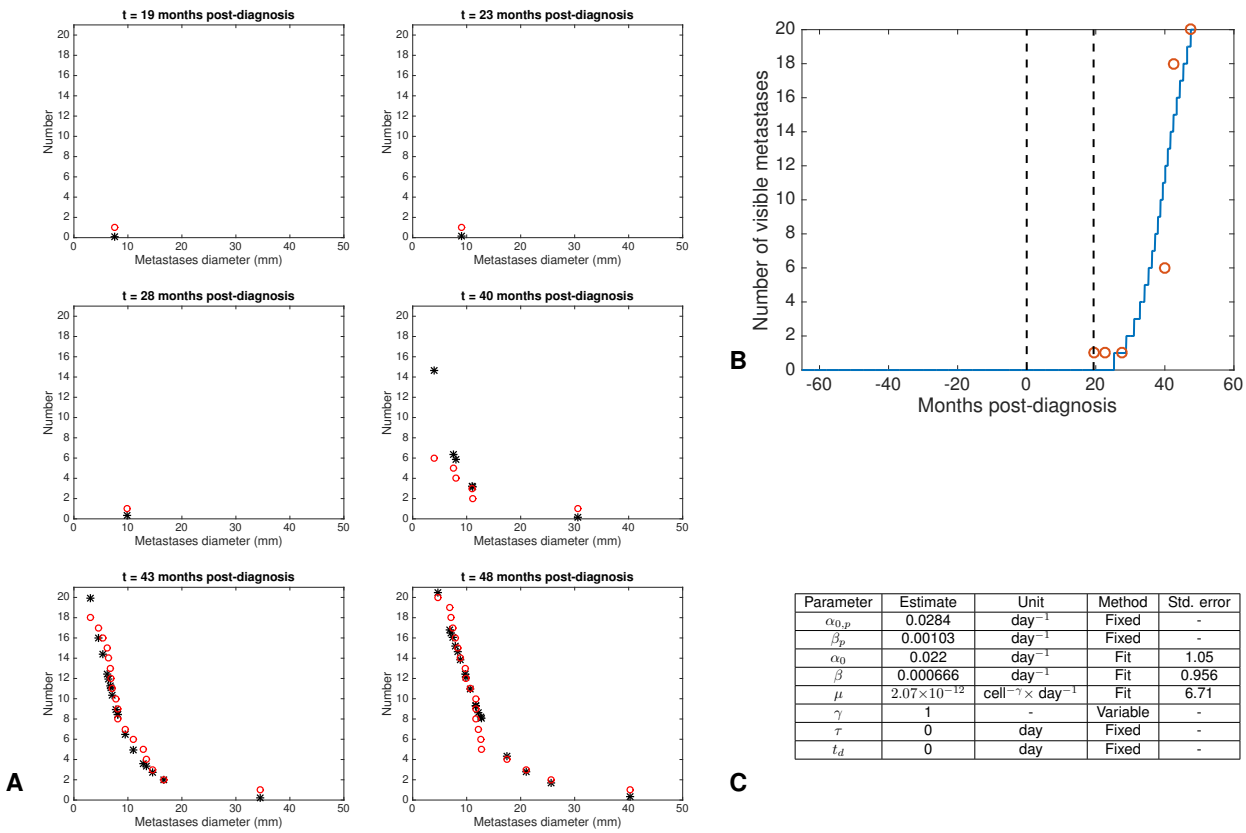
A. Time course of the visible brain metastases (BMs) cumulative size distributions during follow-up. Black stars = model. Red circles = data.

B. Time course of the number of visible BMs.

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

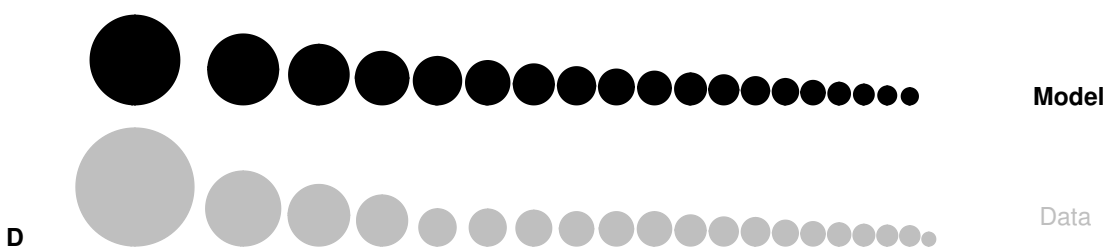
D. Comparison of the BM size distribution between the model fit and the data at the last time point (48 months post-diagnosis).

Figure S4: 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	-



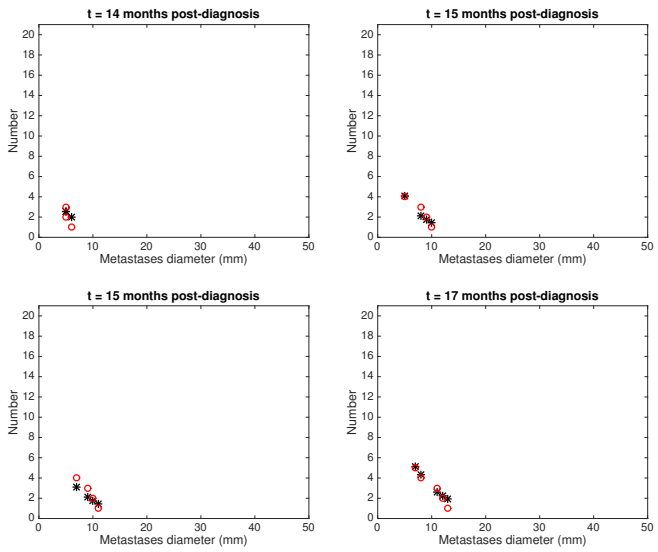
A. Time course of the visible brain metastases (BMs) cumulative size distributions during follow-up. Black stars = model. Red circles = data.

B. Time course of the number of visible BMs.

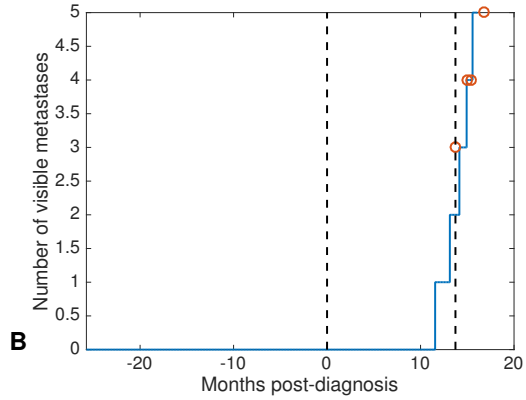
C. Parameter estimates. Std error = Standard errors expressed in percent.

D. Comparison of the BM size distribution between the model fit and the data at the last time point (48 months post-diagnosis).

Figure S5: Fit of the dormancy model for patient 2



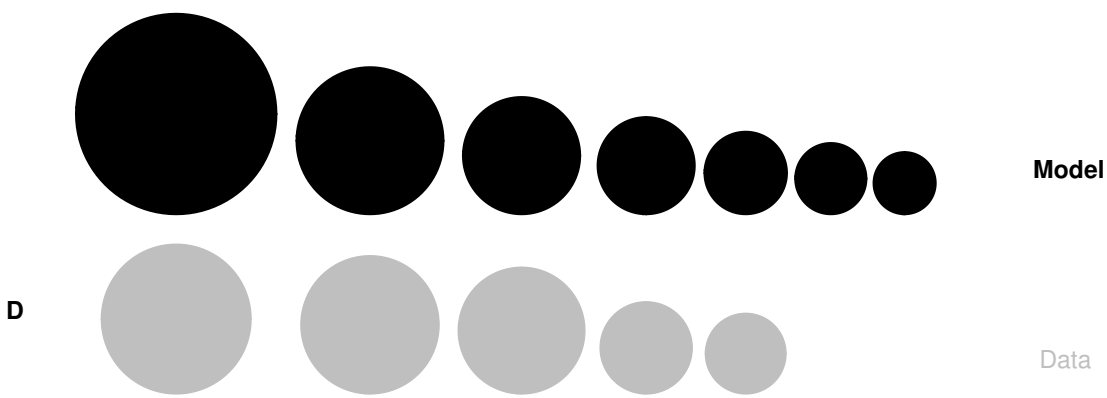
A



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	-

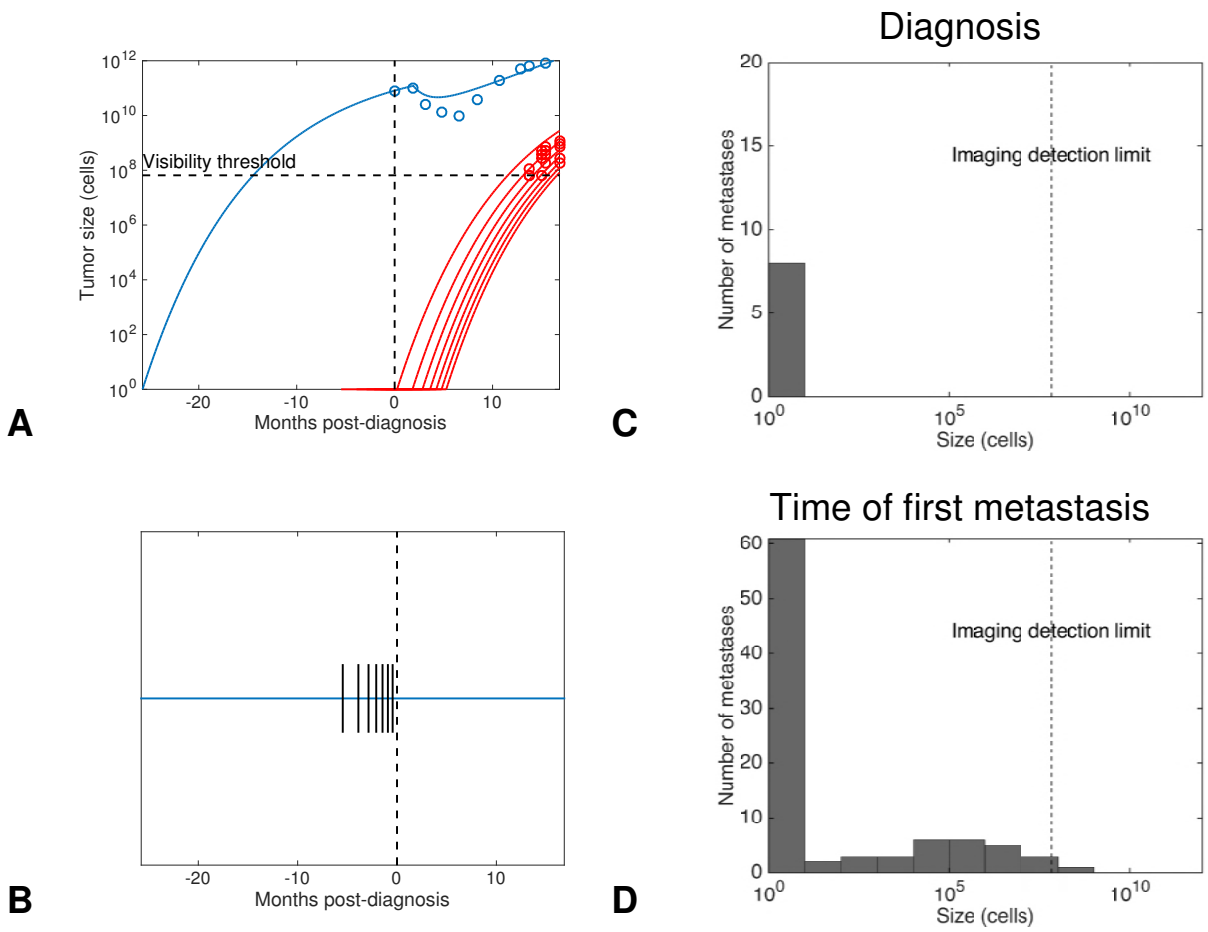
C



D

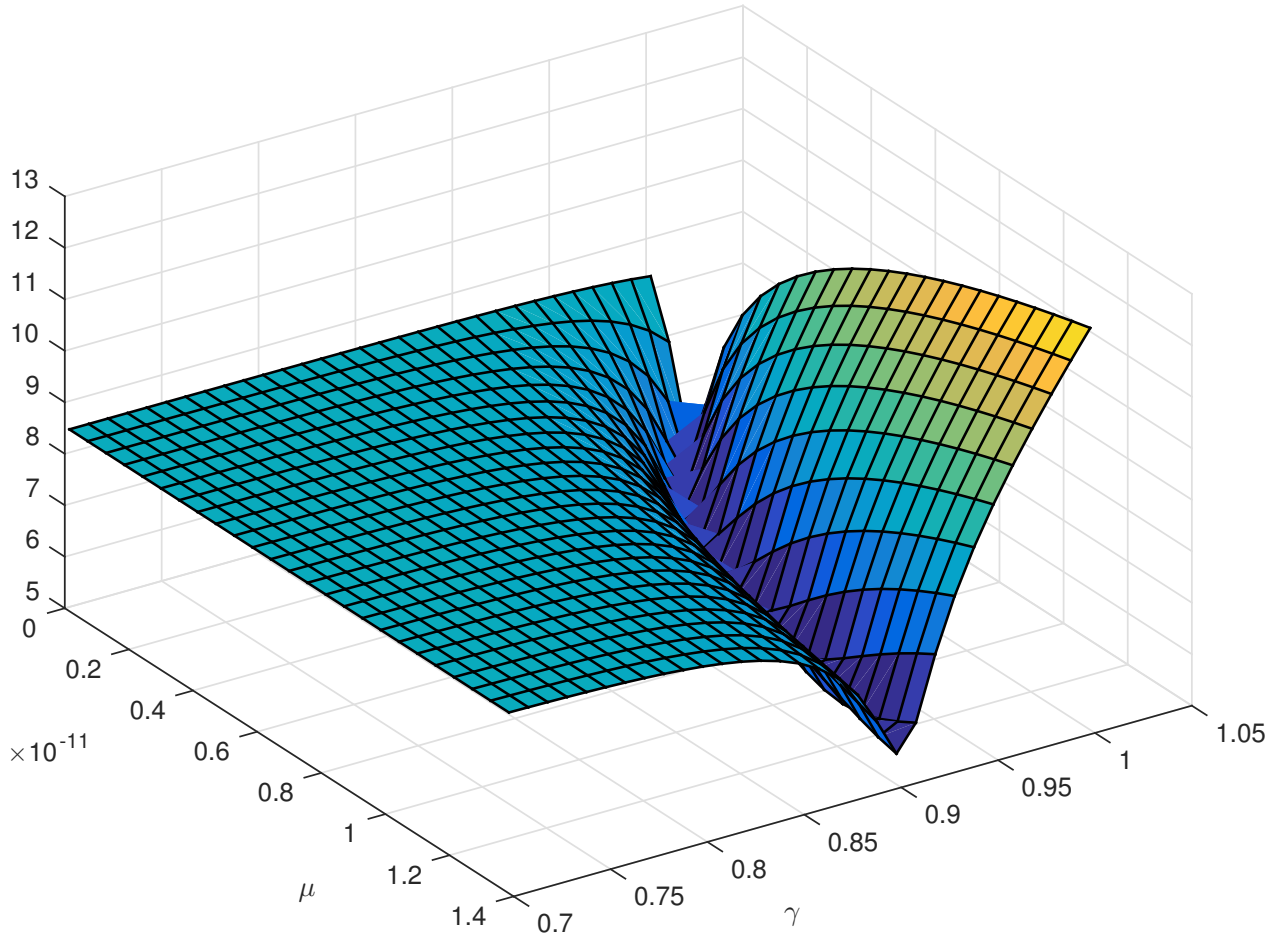
A. Time course of the visible brain metastases (BMs) cumulative size distributions during follow-up. Black stars = model. Red circles = data.
B. Time course of the number of visible BMs.
C. Parameter estimates. Std error = Standard errors expressed in percent.
D. Comparison of the BM size distribution between the model fit and the data at the last time point (48 months post-diagnosis).

Figure S6: Clinical inference for patient 2



A. Inferred growth kinetics of the primary tumor (in blue) and the brain metastases.
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 S7: 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.