



**HAL**  
open science

# **Development and Validation of a Prediction Model of Overall Survival in High-Risk Neuroblastoma Using Mechanistic Modeling of Metastasis**

Sébastien Benzekry, Coline Sentis, Carole Coze, Laëtitia Tessonnier, Nicolas André

## ► **To cite this version:**

Sébastien Benzekry, Coline Sentis, Carole Coze, Laëtitia Tessonnier, Nicolas André. Development and Validation of a Prediction Model of Overall Survival in High-Risk Neuroblastoma Using Mechanistic Modeling of Metastasis. *JCO Clinical Cancer Informatics*, 2021, 5, pp.81-90. <10.1200/CCI.20.00092>. <hal-03147049>

**HAL Id: hal-03147049**

**<https://inria.hal.science/hal-03147049v1>**

Submitted on 19 Feb 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



# Development and Validation of a Prediction Model of Overall Survival in High-Risk Neuroblastoma Using Mechanistic Modeling of Metastasis

Sébastien Benzekry, PhD<sup>1</sup>; Coline Sentis, MD<sup>2</sup>; Carole Coze, MD, PhD<sup>2,3</sup>; Laëtitia Tessonier, MD, PhD<sup>4</sup>; and Nicolas André, MD, PhD<sup>2,5</sup>

abstract

Prognosis of high-risk neuroblastoma (HRNB) remains poor despite multimodal therapies. Better prediction of survival could help to refine patient stratification and better tailor treatments. We established a mechanistic model of metastasis in HRNB relying on two processes: growth and dissemination relying on two patient-specific parameters: the dissemination rate  $\mu$  and the minimal visible lesion size  $S_{vis}$ . This model was calibrated using diagnosis values of primary tumor size, lactate dehydrogenase circulating levels, and the meta-iodobenzylguanidine International Society for Paediatric Oncology European (SIOPEN) score from nuclear imaging, using data from 49 metastatic patients. It was able to describe the data of total tumor mass (lactate dehydrogenase,  $R^2 > 0.99$ ) and number of visible metastases (SIOPEN,  $R^2 = 0.96$ ). A prediction model of overall survival (OS) was then developed using Cox regression. Clinical variables alone were not able to generate a model with sufficient OS prognosis ability ( $P = .507$ ). The parameter  $\mu$  was found to be independent of the clinical variables and positively associated with OS ( $P = .0739$  in multivariable analysis). Critically, addition of this computational biomarker significantly improved prediction of OS with a concordance index increasing from 0.675 (95% CI, 0.663 to 0.688) to 0.733 (95% CI, 0.722 to 0.744,  $P < .0001$ ), resulting in significant OS prognosis ability ( $P = .0422$ ).

JCO Clin Cancer Inform 5:81-90. © 2021 by American Society of Clinical Oncology

## INTRODUCTION

Neuroblastoma is the second most frequent solid tumor in children (8%-10% of children cancers in the United States and Europe) with a median age at diagnosis around 2 years.<sup>1,2</sup> Neuroblastoma is responsible for almost 15% of childhood deaths by cancer.<sup>3</sup> Neuroblastoma is a quite heterogeneous disease at clinical, histological, and biological levels.<sup>4</sup> Consequently, its prognostic spectrum is also wide.<sup>5</sup> The International Neuroblastoma Risk Group (INRG) proposed in 2009 a classification model depending on various factors such as dissemination, histology, grade of tumor differentiation, and genetic abnormalities such as MYCN amplification<sup>6</sup> and age.<sup>3,7</sup> Consequently, neuroblastoma is divided into 3 risk groups: low-, intermediate-, and high-risk neuroblastoma (HRNB), which display different survival rates. For patients treated according to the International Society for Paediatric Oncology European (SIOPEN) recommendations, 5-year overall survival (OS) is more than 90% for the first group, thanks to minimal therapeutics (surgery and/or chemotherapy or simple overseeing); 60% to 80% for the second group,<sup>5</sup> and <

50% for the last group, representing nearly 50% of patients,<sup>3,8-11</sup> despite intensive multimodal treatments. Furthermore, patients progressing during or after initial response to induction have a dismal 5-year event-free survival (< 20% of patient with early progressive disease<sup>12,13</sup>). Although treatment is relatively effective in the low- and intermediate-risk groups, better stratification is needed in HRNB. For these, identifying patients likely to relapse or even progress during therapy may allow avoiding noneffective treatment and explore upfront more experimental therapies.

As early as 1964,<sup>14</sup> efforts have been made to develop mathematical models to assist cancer research.<sup>15</sup> Their aim was to understand multiple biological processes involved in cancer and to propose rational tools for the design of therapeutic drug regimen.<sup>16-18</sup> Three main types of mathematical models can be distinguished. On the one hand, highly complex multiscale models try to integrate as much of the biology as possible, ranging from intracellular molecular processes to systemic interactions at the whole organism level.<sup>19</sup> This approach requires many parameters, and

### ASSOCIATED CONTENT

#### Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on November 5, 2020 and published at [ascopubs.org/journal/cci](https://ascopubs.org/journal/cci) on January 13, 2021: DOI <https://doi.org/10.1200/CCI.20.00092>

consequently, the models are often impossible to reliably calibrate for clinical purposes. On the other hand, purely statistical models and artificial intelligence techniques rely on agnostic algorithms that try to learn patterns directly from the data,<sup>20</sup> with applications mostly in genomics and radiology, but rarely in clinical oncology. In between, mechanistic or semimechanistic models seek to describe only the main determinants of a cancer disease, for a given purpose (eg, understanding<sup>21</sup> or prediction<sup>22</sup> of metastatic relapse).

To our knowledge, no mechanistic model of metastasis has yet been established and validated for clinical neuroblastoma. In this study, we define such a model for HRNB and calibrate it using quantitative data available at diagnosis. We then evaluate its prognostic value in a patient-specific predictive model of OS.

## MATERIALS AND METHODS

The full code and data used to obtain the results are available online.<sup>22a</sup>

### Ethics Statement

Authorization to perform this study was obtained at AP-HM (University Hospitals of Marseille) Health Data Access Portal (number request 32PTJ5). We respect the Informatics and Liberty Law (1978) for the use of data. All parents and patients when appropriate gave consent to participate in this study.

### Data

Our population is made of 49 patients with HRNB (see the stratification algorithm in the Data Supplement) treated according to the HRNBL1 protocol recommendations (Data Supplement)<sup>9</sup> in the pediatric hematology and oncology unit of AP-HM between November 26, 2007, and August 30, 2018. For survival analyses, end date was either the date of the patient's death or the date of last news. Inclusion criteria were the ones of the HRNBL1 protocol<sup>9</sup> (Data Supplement).

The data collected at diagnosis consisted of six variables: age, sex, MYCN amplification status, tumor size, lactate dehydrogenase (LDH) level—researched by polymerase chain reaction from peripheral blood and/or from primary or metastatic tumor tissue—and SIOOPEN score. The data set is available in the Data Supplement.

### SIOOPEN Scoring

Meta-iodobenzylguanidine (MIBG) is known to bind to neuroblastoma cells using iodine 123 (I123),<sup>23</sup> and MIBG scintigraphy is consequently used to evaluate the extent of the disease.<sup>24,25</sup> We used the semiquantitative SIOOPEN score that was elaborated to predict extension and severity of the disease.<sup>26</sup> Details about its definition are given in the Data Supplement. A high score has been shown as

pejorative, but no reproducible cutoff has yet been found.<sup>24,26</sup>

### Tumor Characteristics

Location and size of primary tumors (PTs) were evaluated using radiological reports performed at diagnosis (Data Supplement). Forty-five patients (91.8%) had undergone computed tomography scanning, and 11 patients (22.4%) had undergone magnetic resonance imaging. PT volumes were estimated using the formula:  $\frac{4}{3}\pi abc$ , with a, b, and c being half the largest, medium, and smallest axes of an ellipsoid tumor, respectively. In case only two measurements were available, the smallest one was used as third.

Location and number of visible metastases were retrieved from MIBG scintigraphy. In addition, bone marrow metastases were searched by performing myelograms and bone marrow biopsies.

Date of progression was recorded as the date on which unfavorable evolution of the disease was highlighted by radiology and nuclear imaging (positron emission tomography and/or MIBG I123 scintigraphy).

### Mathematical Model

**Definition.** The mathematical model was based on a previously published framework,<sup>21,27</sup> which allows us to simulate a cancer disease at the organism scale, including growth of the PT and birth and growth of secondary lesions (Fig 1). We assumed growth of both the primary and secondary tumors to follow an exponential law,

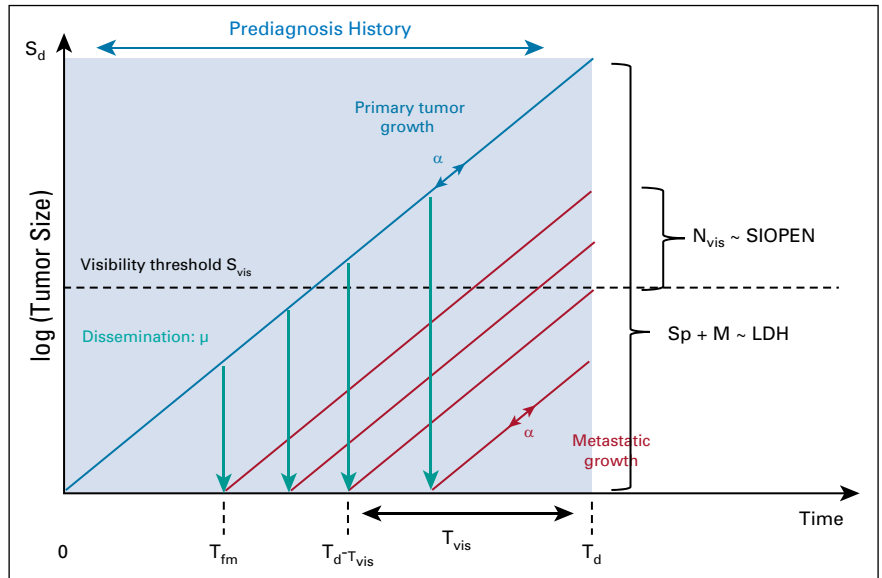
$$S_p(t) = S(t) = e^{\alpha t},$$

where  $S_p(t)$  and  $S(t)$  denote the sizes of a primary and a secondary tumor (expressed in number of cells), starting from one cell at time  $t = 0$ . This differs from previous works where we used Gompertzian growth<sup>21,27</sup> and is due to the pediatric aspect of the tumors considered here. Indeed, we do not have evidence of growth slowdown in this context and tumor age estimates resulting from the exponential assumption were biologically plausible, in contrast to other cases (eg, adult lung tumors<sup>21</sup>). The parameter  $\alpha$  denotes the proliferation rate. Assuming a metastasis birth rate proportional to the PT size with parameter  $\mu$ , the number of metastases at time  $t$  is given by<sup>21</sup>

$$N(t) = \mu \int_0^t S_p(s) ds = \frac{\mu}{\alpha} (e^{\alpha t} - 1).$$

The parameter  $\mu$  corresponds to the per day probability for each cell of the PT to spread and establish a distant metastasis. The total metastatic burden (total number of metastatic cells in the organism) is given by<sup>27</sup>

**FIG 1.** Schematic of the mathematical model. Primary and secondary tumors are assumed to have exponential growth kinetics governed by a proliferation rate  $\alpha$ . Dissemination of metastasis is controlled by the parameter  $\mu$ . From these and PT size at diagnosis  $S_d$ , the PT age  $T_d$  can be computed and simulations of the natural history can be performed. Adjunction of a visibility threshold  $S_{vis}$  results in predictions of the number of visible metastases  $N_{vis}$  and total cancer mass (primary plus secondary tumors)  $S_p$  plus  $M$ . These are respectively compared with the SIOPEN score and lactate dehydrogenase level. The time of birth of the first metastasis is denoted  $T_{fm}$  and the time to reach  $S_{vis}$  from one cell  $\tau_{vis}$ . Note that the number of visible metastases at time  $T_d$  is the total number of metastases at time  $T_d - \tau_{vis}$ . LDH, lactate dehydrogenase; PT, primary tumor; SIOPEN, International Society for Pediatric Oncology European.



$$M(t) = \mu \int_0^t S_p(s) S(t-s) ds = \mu t e^{\alpha t}.$$

Visible metastases at time  $t$  (ie, metastases with size larger than a visibility threshold  $S_{vis}$ ) are the ones that were born early enough to have reached  $S_{vis}$  at  $t$ , that is, before  $t - \tau_{vis}$ , where  $\tau_{vis}$  is the time to reach  $S_{vis}$  (Fig 1). This time is given by  $\tau_{vis} = \frac{\ln(S_{vis})}{\alpha}$ , and the number of visible metastases can then be computed as follows:

$$N_{vis}(t) = N(t - \tau_{vis}) = \mu \int_0^{t - \tau_{vis}} S_p(s) ds = \frac{\mu}{\alpha} \left( \frac{e^{\alpha t}}{S_{vis}} - 1 \right).$$

The visibility threshold  $S_{vis}$  is considered a patient-specific model parameter.

For forward simulations of the model, a discrete version was employed with initiation time  $T_j$  and size  $S_j$  of the  $j$ -th metastasis given by

$$\begin{cases} N(T_j) = j \Leftrightarrow T_j = \frac{1}{\alpha} \ln \left( j \frac{\alpha}{\mu} - 1 \right) \\ S_j(t) = e^{\alpha(t - T_j)}, \quad \text{for } t > T_j \end{cases}$$

**Parameter calibration.** Leveraging information available in the literature, the parameter  $\alpha$  was assumed to be patient-nonspecific. We searched PubMed for studies related to the cell doubling time (CDT) in neuroblastoma and CDTs from a database of commercial cell population (Data Supplement). All cell lines were obtained from human patients, and CDTs were established in vitro. Of the 73

strains studied, 15 were excluded because of a lack of knowledge regarding possible exposure to chemotherapy. The median CDT was 48 hours (20-252 hours). We then fixed  $\alpha = \frac{\ln 2}{\text{CDT}_{\text{median}}}$ . Denoting with superscript  $i$  the quantities that depend on individual  $i$ , this further allowed us to compute an estimation of the PT age (or time of diagnosis,  $T_d^i$ ) from the PT size data  $S_d^i$ ,

$$T_d^i = \frac{\ln(S_d^i)}{\alpha}.$$

The volume  $S_d^i$  was converted into a number of cells using the assumption  $1 \text{ mm}^3 \approx 10^6$  cells.<sup>28</sup>

For each patient, two quantitative measurements were used to compare the metastatic model with the data: the SIOPEN score and the LDH blood level. The former was assumed to be a surrogate of the visible number of metastases and the latter to represent the total cancer burden in the organism (PT plus metastases, Fig 1). We thus assumed

$$\text{SIOPEN}^i = N_{vis}(T_d^i; \mu^i, S_{vis}^i) \times (1 + \sigma \epsilon^1), \quad \epsilon^1 \sim N(0, 1),$$

$$\text{LDH}^i = \phi \left( S_p(T_d^i) + M(T_d^i; \mu^i, S_{vis}^i) \right) \times (1 + \sigma \epsilon^2), \quad \epsilon^2 \sim N(0, 1),$$

which express a proportional error model for the observations with standard deviation  $\sigma = 0.1$ , corresponding to a 10% measurement error. The parameter  $\phi$  represents a patient-nonspecific conversion coefficient between the number of cells and LDH (expressed in international units per liter, IU/L). Maximization of the log-likelihood for the above expression leads to minimization of the following objective function:

$$\begin{aligned}
l(\mu^i, S_{vis}^i, \phi) &= l_{SIOPEN}(\mu^i, S_{vis}^i) + l_{LDH}(\mu^i, \phi) \\
&= \frac{(SIOPEN^i - SIOPEN_{model}^i(\mu^i, S_{vis}^i))^2}{2(\sigma SIOPEN_{model}^i(\mu^i, S_{vis}^i))^2} \\
&\quad + \ln \left( \sigma \sqrt{2\pi} SIOPEN_{model}^i(\mu^i, S_{vis}^i) \right) \\
&\quad + \frac{(LDH^i - LDH_{model}^i(\mu^i, \phi))^2}{2(\sigma LDH_{model}^i(\mu^i, \phi))^2} \\
&\quad + \ln \left( \sigma \sqrt{2\pi} LDH_{model}^i(\mu^i, \phi) \right),
\end{aligned}$$

with

$$\begin{aligned}
SIOPEN_{model}^i(\mu^i, S_{vis}^i) &= N_{vis}(T_d^i; \mu^i, S_{vis}^i) = \frac{\mu^i}{\alpha} \left( \frac{S_d^i}{S_{vis}^i} - 1 \right), \\
LDH_{model}^i(\mu^i, \phi) &= \phi \left( S_p(T_d^i) + M(T_d^i; \mu^i, S_{vis}^i) \right) \\
&= \phi \left( S_d^i + \frac{\mu^i}{\alpha} S_d^i \ln \left( \frac{S_d^i}{S_{vis}^i} \right) \right).
\end{aligned}$$

The explicit formula can be derived from expressions (2) and (3). The parameter  $\phi$  was arbitrarily fixed to  $10^{-9}$  IU/L/cell from preliminary simulations. Minimization was implemented using the Nelder-Mead algorithm of the *minimize* function of the *scipy* python package (version 3.7).

### Statistical Methods for Predictive Analysis

Because of ranges spanning several orders of magnitude, individual values of LDH levels and the mathematical parameter  $\mu$  were log-transformed beforehand. Correlations between variables were computed using the *nominal.associations* function of the *dython python* package, that is, using Pearson's correlation, correlation ratio, and Cramér's V for continuous-continuous, categorical-continuous, and categorical-categorical associations, respectively. Associations between variables and progression-free or OS were assessed using log-rank tests for dichotomized groups and univariate and multivariable proportional hazard Cox regression models. The *lifelines* python package was used to fit the models.<sup>29</sup> Resulting models were evaluated for their predictive power by computing the mean of Harrell's c-index<sup>30</sup> during 100 replicates of a five-fold cross-validation procedure. Specifically, we first selected the variables below a  $P = .2$  significance threshold in multivariable analysis. Then, multivariable Cox models including only these

variables were trained on the cross-validation learning sets, and the c-index was computed on the corresponding test sets. This procedure was repeated 100 times with randomly selected partitions of the data set to assess variability of the results with respect to data sampling. For construction of the prognosis score, the Cox model was trained using the selected variables on the entire data set. For a patient  $i$  with covariates  $x^i$ , the score was defined by  $\beta^T x^i$ , with  $\beta$  being the vector of the Cox coefficients. Calibration of these Cox models was assessed using calibration plots at the median survival value, using the *calibrate* function of the *rms R* package.

## RESULTS

Description of the cohort is provided in Table 1 and the Data Supplement. Overall survival and progression-free survival are plotted in the Data Supplement.

### Descriptive Power of the Mathematical Model

To describe the metastatic burden of patients with HRNB, we developed a semimechanistic modeling approach whereby the metastatic process is reduced to two main phenomena: growth and dissemination (Fig 1). Growth was assumed to be exponential with growth rate  $\alpha$  (day<sup>-1</sup>) and dissemination to be proportional to the PT size, with a proportionality factor  $\mu$  (metastases·cell<sup>-1</sup>·day<sup>-1</sup>). In addition, the model depends on the minimal visible size for MIBG detection  $S_{vis}$  (mm), a coefficient  $\phi$  converting LDH values in the number of cells (IU·L<sup>-1</sup>·cell<sup>-1</sup>), and the age

TABLE 1. Patient Characteristics

Sex	
Male	n = 28 (57.2%)
Female	n = 21 (42.8%)
Age	
≥ 18 mo	n = 43 (87.8%)
< 18 mo	n = 6 (12.2%)
LDH rates	
≥ 1250 UI/L	n = 18 (36.7%)
< 1250 UI/L	n = 31 (63.3%)
MYCN	
Amplified	n = 23 (46.9%)
Nonamplified	n = 26 (53.1%)
SIOPEN	
≥ 4	n = 30 (65.2%)
Only if MIBG was positive < 4	n = 16 (34.8%)
Metastasis	
Present	n = 43 (87.8%)
Absent	n = 6 (12.2%)

Abbreviations: LDH, lactate dehydrogenase; MIBG, meta-iodobenzylguanidine; MYCN, v-myc avian myelocytomatosis viral related oncogene, neuroblastoma derived; SIOPEN, International Society for Pediatric Oncology European.

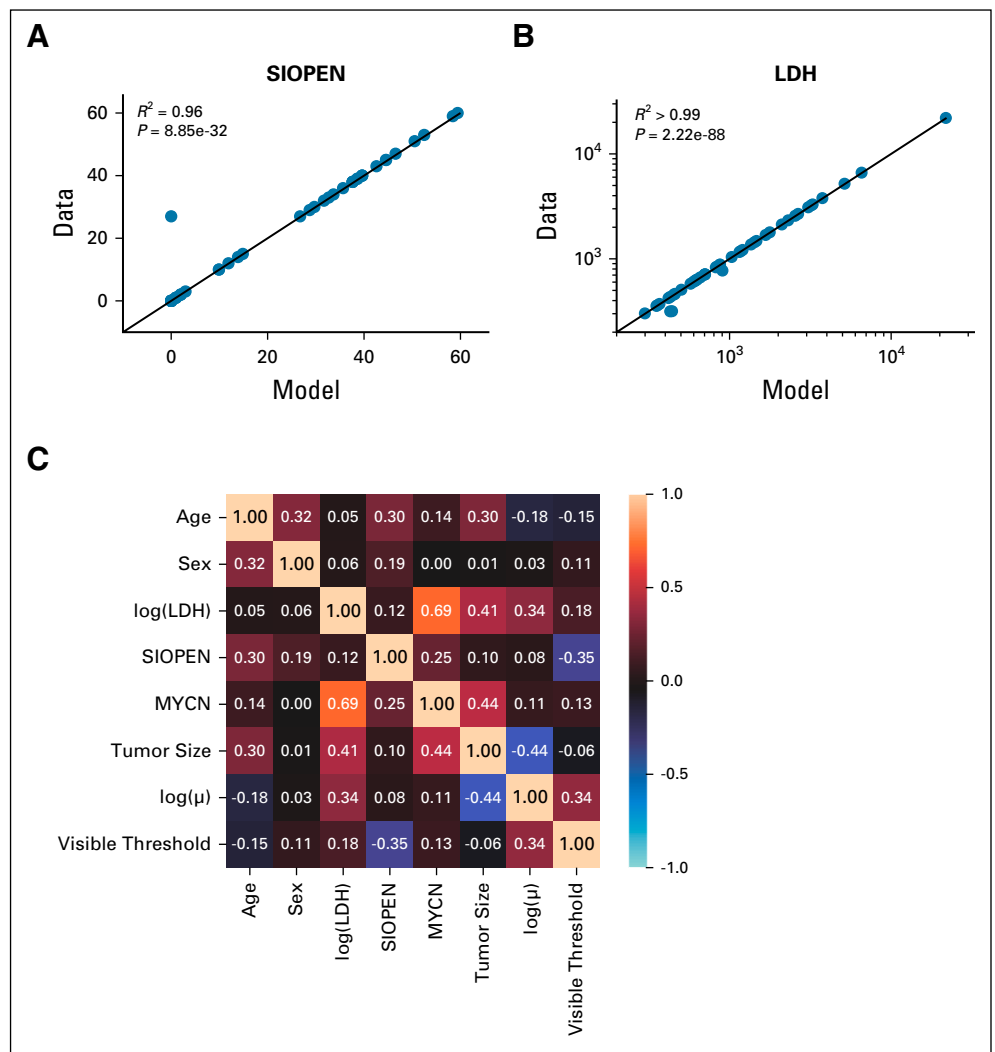
of the PT at diagnosis  $T_d$ . To estimate these 5 parameters, the only quantitative data available were the PT size at diagnosis  $S_d$ , the LDH, and the SIOPEN. To avoid overfitting, restrictive assumptions had to be made:  $\alpha$  and  $\phi$  were considered constant across the patients and fixed from the literature and exploratory simulations, respectively;  $T_d^i$  was determined from  $S_d^i$  in patient  $i$ , and  $\mu^i$  and  $S_{vis}^i$  were patient-specific and determined from the LDH and SIOPEN values. Initial attempts considering linking the SIOPEN to the visible metastatic mass (instead of number) or  $S_{vis}^i = S_{vis} \forall i$  and only  $\mu^i$  patient-specific were unable to describe the data.

We assumed that the LDH level was a surrogate of the total tumor mass (primary plus secondary),<sup>6,9,31,32</sup> whereas the SIOPEN score reflected the number of visible metastases (Fig 1). The model was able to accurately reproduce the SIOPEN score and LDH levels (Figs 2A and 2B,  $R^2 = 0.96$  and  $> 0.99$ , respectively). Interestingly, the parameter  $\ln \mu$  revealed no correlation with either the  $\log(\text{LDH})$  ( $R = 0.25$ ) or the

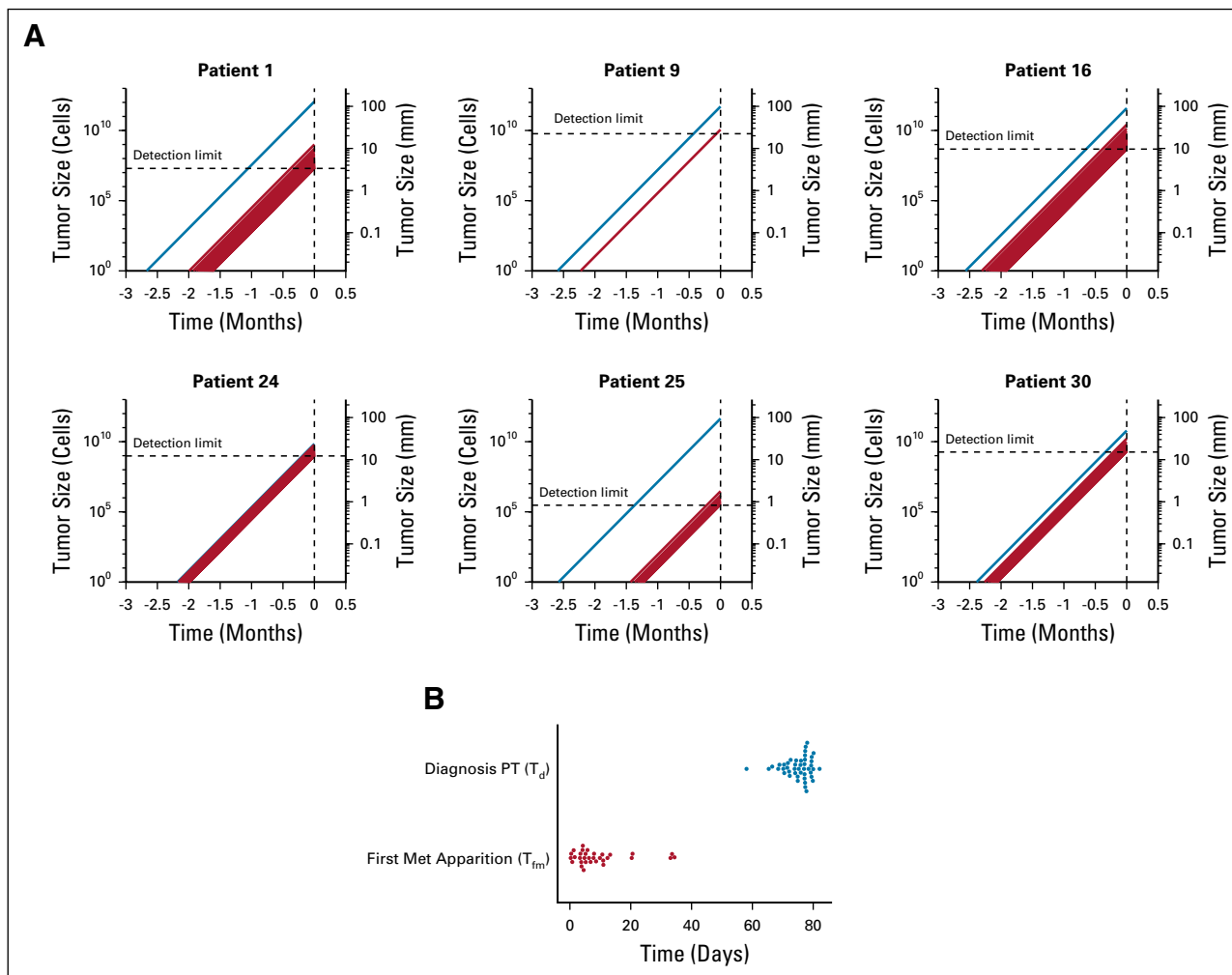
SIOPEN ( $R = 0.201$ ), but some correlation with the tumor size ( $R = -0.44$ ), see Figure 2C.

### Simulations of the Natural History of HRNB

Statistical inference of the model parameters  $\mu^i$  and  $S_{vis}^i$  allowed us to perform simulations of the predicted natural history of the disease. Representative patients are shown in Figure 3A (see the Data Supplement for all patients). These highlight the high interindividual variability, well captured by the mathematical model. As a general observation, however, we predicted that once initiated, metastatic dissemination then experienced a burst, with multiple metastasis born in a short period of time (see patients 1, 16, 24, 25, and 30 in Fig 3A). Nevertheless, the time  $T_{fm}$  between the first cancer cell and birth of metastasis was variable, from almost simultaneous to cancer initiation (eg, patient 24, 0.42 days) to one month (eg, patient 25, 34.4 days), see Figure 3B. The predicted age of the PT was less variable (median, 76 days, range, 58.1-82.1 days).



**FIG 2.** Descriptive power of the mathematical model. (A) Fit of the SIOPEN data. The solid line is the identity line. (B) Fit of the LDH data. The solid line is the identity line. (C) Correlation matrix of all features including clinical variables and (log) of the mathematical parameter  $\mu$ . LDH, lactate dehydrogenase; MYCN, v-myc avian myelocytomatosis viral related oncogene, neuroblastoma derived; SIOPEN, International Society for Pediatric Oncology European.



**FIG 3.** Mechanistic simulations of the prediagnosis history of patients with high-risk neuroblastoma. A mechanistic model was calibrated using patient data of SIOOPEN score (number of visible metastases) and LDH levels (total cancer burden), which resulted in individual values of the model parameters  $\mu$  and  $S_{vis}$ . These values were used to simulate the prediagnosis natural history of the disease. (A) Simulations of the PT (blue) and metastasis (red) growth kinetics, for representative patients. (B) Distributions of the times of the PT diagnosis (age of the PT,  $T_d$ ) and birth of the first metastasis ( $T_{fm}$ ). LDH, lactate dehydrogenase; PT, primary tumor; SIOOPEN, International Society for Pediatric Oncology European.

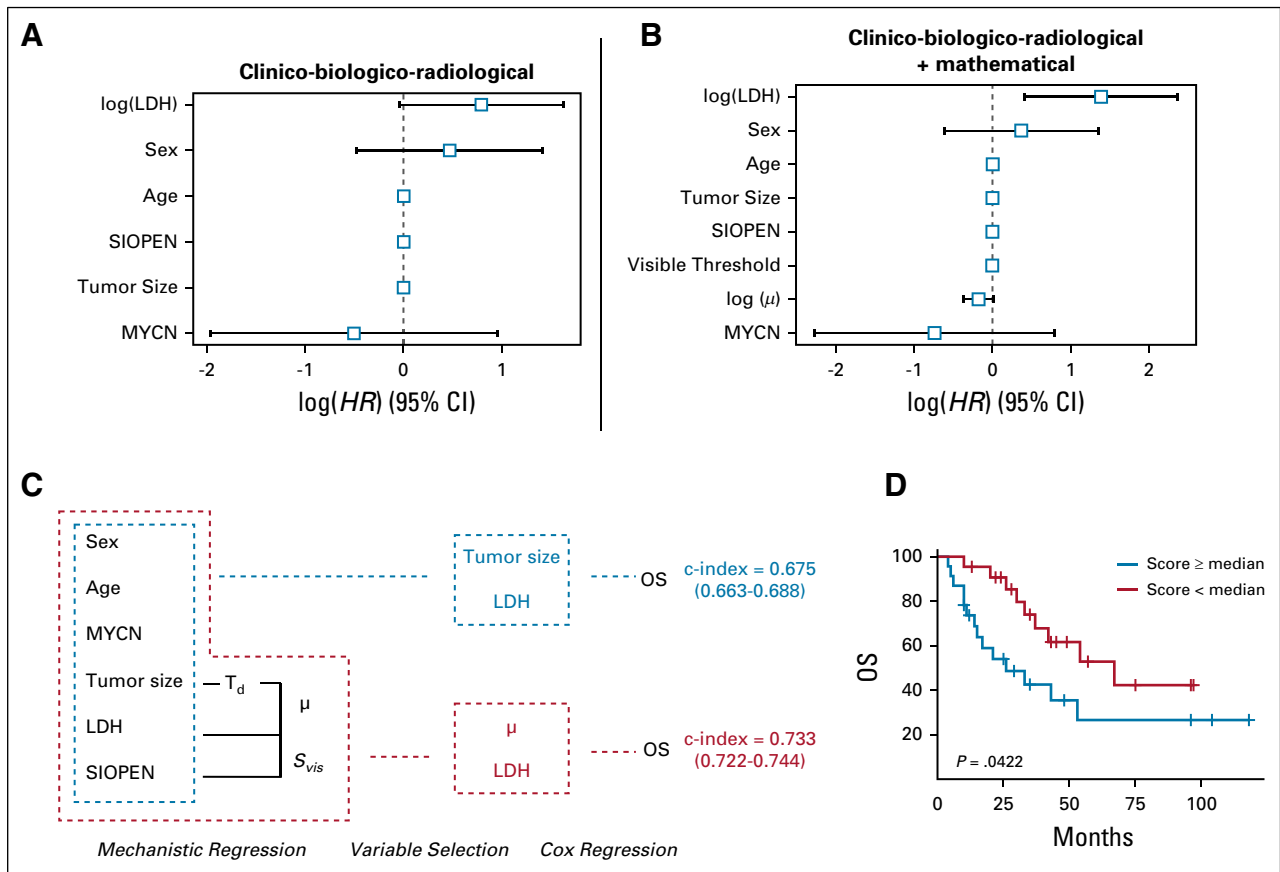
### Prognostic Analysis

**Analysis of prognosis factors.** Using log-rank tests for dichotomized groups (Data Supplement), no significant difference was found in OS for sex ( $P = .198$ ), MYCN status ( $P = .181$ ), age ( $P = .757$ ), or tumor size ( $P = .147$ ). A significant difference in OS was found only at the 80th and 90th percentiles of LDH levels and SIOOPEN score. Using univariate Cox regression, LDH ( $P = .00794$ ) and tumor size ( $P = .00621$ ) were significantly associated with OS, although only LDH remained in multivariable analysis (Fig 4A). The computational biomarker  $\mu$  seemed to be the quantitative parameter most associated with OS in dichotomized analysis (Data Supplement), even if not statistically significant probably due to a crossing of the survival curves. Confirming this observation,  $\mu$  was independently associated with better OS in multivariable Cox regression analysis (HR = 0.837 [95% CI, 0.689 to 1.02],  $P = .0739$ , Table 2 and Fig 4B). On

the other hand, consistent with its definition linked to variability being caused by random factors affecting the detection of metastases at MIBG imaging,  $S_{vis}$  was not associated with OS ( $P = .549$ ). For progression-free survival (PFS), no variable demonstrated significant association (Data Supplement).

### Prediction models and added value of the mechanistic model.

To assess the utility of the mechanistic approach, we compared two Cox regression-based prediction models: either using the clinico-biologico-radiological data alone or augmenting them with the computational biomarkers  $\mu$  and  $S_{vis}$  computed from regression of the mechanistic model (Fig 4C). Initial prediction models integrating all the variables demonstrated very poor predictive abilities in both cases (c-index = 0.592 [95% CI, 0.578 to 0.606] and 0.582 [0.567 to 0.597], respectively). Consequently, variable selection was performed by selecting only those variables



**FIG 4.** Prognosis of overall survival. Using multivariable Cox regression, we compared the association and prognostic value of clinical variables alone (left) or complemented with the mathematical biomarkers (right). (A) Hazard ratios and 95% CIs of the clinical variables in multivariable Cox regression. The reported c-index corresponds to the one obtained in cross-validation using variables selected at  $P < .2$  in multivariable Cox analysis (log(LDH)). (B) Separation of patients according to the Cox score predicted from the selected clinical variable (log(LDH)). (C) Schematic representation of the OS predictive analyses. In blue, using only the clinical variables and Cox regression. In red, using in addition the computational biomarkers derived from mechanistic modeling. The analysis was composed of three phases: 1) mechanistic regression for estimation of the computational biomarkers from the data, 2) variable selection and 3) final Cox regression model. (D) Stratification of the patients based on the final model including the computational biomarker  $\mu$ . The  $P$ -value is from a log-rank test. LDH, lactate dehydrogenase; OS, overall survival; PT, primary tumor.

with  $P < .2$  in multivariable regression. This selected tumor size and log(LDH) in the clinical model and log( $\mu$ ) and log(LDH) in the clinical plus mathematical model. The predictive power was improved to a c-index of 0.675 (0.663-0.688) for the clinical model. Critically, the clinical plus mathematical model demonstrated a substantially and significantly larger c-index of 0.733 (0.722-0.744,  $P < .0001$ , one-tailed  $t$ -test), indicating predictive benefit of the mechanistic approach. This was also confirmed by better calibration plots (Data Supplement) and allowed us to define a stratification of patients with HRNB based on the Cox score with significant discriminatory ability ( $P = .0422$ , Fig 4D).

## DISCUSSION

In mathematical oncology, few studies have been devoted to modeling the metastatic process and even fewer have used quantitative clinical data for their calibration.<sup>21,22,33-37</sup> Among these, almost none has been translated into a predictive

tool,<sup>22</sup> and to our knowledge, this is the first report of such a study for neuroblastoma. We used a mechanistic approach based on classical risk factors, routinely collected at diagnosis by clinicians. Our model could adequately describe total cancer mass represented by LDH levels and visible metastases represented by the SIOPEN scores. Beyond this mere descriptive value, our analysis suggested significant added value when including the computational biomarker  $\mu$  into a patient-specific prognostic model of OS.

Tumor growth and metastatic progression are complex biological processes that include tumor proliferation, regulation of abnormalities concerning stem cells,<sup>38</sup> neoangiogenesis,<sup>39</sup> microenvironment interactions,<sup>38,39</sup> immune interactions with tumor cells, and dysregulation immune system.<sup>4,39,40</sup> These complex interacting processes are regulated by many genes or epigenetic regulators<sup>41</sup> currently still being investigated. How to model these complex properties remains an open debate. Here, we used a semimechanistic approach relying on a simplified description of these phenomena, because of

**TABLE 2.** Multivariable Cox Analysis of Overall Survival

Covariate	Hazard ratio	P	Lower 95%	Upper 95%
Age	1	0.722	0.985	1.02
Sex	1.44	0.463	0.54	3.86
log(LDH)	4.01	0.00562	1.5	10.7
SIOPEN	0.999	0.963	0.966	1.03
MYCN	0.477	0.346	0.102	2.22
Tumor size	1	0.99	0.999	1
log( $\mu$ )	0.837	0.0739	0.689	1.02
Visible threshold	0.997	0.549	0.988	1.01

Abbreviations: LDH, lactate dehydrogenase; MYCN, v-myc avian myelocytomatosis viral related oncogene, neuroblastoma derived; SIOPEN, International Society for Pediatric Oncology European.

the limited data available from clinical and radiological observations. To limit overfitting, very restrictive assumptions had to be made, including a patient-nonspecific value of  $\alpha$  (doubling time) despite potentially important interindividual variability, the same growth law and parameter for primary and secondary tumors, or an extreme simplification of the metastatic process. Note that from expressions (4) and (5), only the ratio  $\frac{\alpha}{\mu}$  is identifiable from the data under our assumptions, thus preventing these parameters to be both patient-specific. Consistent with the aggressive nature of this disease, dissemination was characterized by a very high value of  $\mu$  (median = 0.0446 metastases/cell/day, range,  $2.3 \times 10^{-6}$ –41) in comparison to other adult cancers (eg, typical  $\mu = 2.30 \times 10^{-12}$  metastases/cell/day in breast cancer). Notably, the inferred values of metastatic sizes at diagnosis and threshold values ( $S_{vis}$ ) individually fall within biologically realistic ranges (median,  $S_{vis}$  10.5 mm, range, 0.65–50 when ignoring the patients with predicted SIOPEN equal to zero, for which  $S_{vis}$  is not identifiable and was estimated to unrealistically large values [see file S1], although no a priori metastatic size information was used when calibrating the model. We attribute the remaining interpatient variability in  $S_{vis}$  to the reading process where scintigraphy images can be quite diffuse and metastases are hard to detect. In contrast, other predictions of the model and limiting assumptions such as the very large total number of metastases and the same doubling time for all patients warrant against a strict mechanistic interpretation of the model.

The main result of this work resides in the statistically significant predictive added value that the computational biomarker  $\mu$  brought over the mere clinical data. Paradoxically, a high  $\mu$  value was a factor of better OS in our

cohort. This might be explained by two possible hypotheses. First, patients with high  $\mu$  may have an aggressive neuroblastoma, which may result in a better sensitivity to chemotherapy and therefore better survival. Second, patients with high  $\mu$  have a bigger total cancer burden, but this total mass could result in systemic inhibition of proliferation.<sup>42</sup> This would be consistent with the fact that  $\mu$  is a good prognosis factor for OS but not for PFS. Indeed, the visible mass might progress while suppressing the growth of smaller invisible tumors. Assuming further that death results from the total mass present in the organism and not only the visible lesions could possibly explain why patients who progress are distinct from patients with the largest number of metastases (ie, largest  $\mu$ ). To further confirm or invalidate these hypotheses, further mechanistic insights could be gained by linking  $\mu$  to molecular analyses of the tumor. The microenvironment and more specifically the immune system might also be implicated in slow tumor progression and a host's tumor long-term control.

Although not being the mainstay yet, individual molecular profiling has been studied in neuroblastoma.<sup>3,4</sup> Several studies explored genome wide associations to predict outcomes for patients<sup>43,44</sup> with HRNB, but they are not used yet in day-to-day clinical practice. The relevance of the identified markers can also be questioned because of the lack of evidence of a causal relationship.<sup>45</sup> To allow using our model for each individual patient, such omics data could be integrated in future work, using more advanced machine learning techniques to deal with the high-dimensional nature of such data.

The major limitation of our study is the limited ( $n = 45$ ) number of patients because of the monocentric nature of our cohort. This nevertheless corresponds to all patients with HRNB treated at our institution over a 10-year period. This prevented us from extracting a test set from the data before any analysis, on which to evaluate the predictive power of the model a posteriori. Nevertheless, we tested the predictive abilities of  $\mu$  using cross-validation, that is, on test sets independent of the learning ones. In this cross-validation process, the model changes on each fold, and thus, deciding on the best model still relies on the entire data set. Critically, the mean c-index was found to be higher when using  $\mu$  to predict survival. Further research should evaluate the predictive value of this final model on independent data.

## AFFILIATIONS

<sup>1</sup>MONC Team, Inria Bordeaux Sud-Ouest and Institut de Mathématiques de Bordeaux, CNRS, University of Bordeaux, Bordeaux, France

<sup>2</sup>Paediatric Hematology and Oncology Department, Hôpital pour enfant de La Timone, AP-HM, Marseille, France

<sup>3</sup>Aix Marseille University, Marseille, France

<sup>4</sup>Department of Nuclear Medicine, Hôpital de La Timone, AP-HM, Marseille, France

<sup>5</sup>SMARTc Unit, Centre de Recherche en Cancérologie de Marseille, Inserm U1068, Aix Marseille University, Marseille, France

**PREPRINT VERSION**

Presented at the medRxiv preprint server: Benzekry S, Sentis C, Coze C, et al: Descriptive and prognostic value of a computational model of metastasis in high-risk neuroblastoma. medRxiv 2020. <http://doi.org/10.1101/2020.03.26.20042192>.

**CORRESPONDING AUTHOR**

Sébastien Benzekry, PhD, Inria Bordeaux Sud-Ouest and Institut de Mathématiques de Bordeaux, 200, avenue de la Vieille Tour, Talence, France; e-mail: [sebastien.benzekry@inria.fr](mailto:sebastien.benzekry@inria.fr).

**ETHICAL APPROVAL AND CONSENT TO PARTICIPATE**

Authorization to perform this study was obtained at AP-HM (University Hospitals of Marseille) Health Data Access Portal (number request 32PTJ5). We respect the Informatics and Liberty Law (1978) for the use of data. All parents and patients when appropriate gave consent to participate in this study.

**EQUAL CONTRIBUTION**

S.B. and C.S. contributed equally to this work.

**SUPPORT**

S. B. is funded by ITMO Cancer AVIESAN and French Institute National du Cancer (grant #19CM148-00). N. André wishes to thank Canceropole PACA, LN la vie, les copains de Charles for their support.

**AUTHOR CONTRIBUTIONS**

**Conception and design:** Sébastien Benzekry, Coline Sentis, Nicolas André

**Administrative support:** Nicolas André

**Provision of study materials or patients:** Carole Coze, Laëtitia Tessonnier, Nicolas André

**Collection and assembly of data:** Coline Sentis, Carole Coze, Laëtitia Tessonnier, Nicolas André

**Data analysis and interpretation:** Sébastien Benzekry, Coline Sentis, Nicolas André

**Manuscript writing:** All authors

**Final approval of manuscript:** All authors

**Accountable for all aspects of the work:** All authors

**AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST**

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to [www.asco.org/rwc](http://www.asco.org/rwc) or [ascopubs.org/cci/author-center](http://ascopubs.org/cci/author-center).

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](http://OpenPayments)).

**Nicolas André**

**Research Funding:** Bristol-Myers Squibb

**Travel, Accommodations, Expenses:** Bristol-Myers Squibb

No other potential conflicts of interest were reported.

**REFERENCES**

1. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2020. *Cancer J Clin* 70:7-30, 2020
2. Ahmed AA, Zhang L, Reddivalla N, et al: Neuroblastoma in children: Update on clinicopathologic and genetic prognostic factors. *Pediatr Hematol Oncol* 34:165-185, 2017
3. Cohn SL, Pearson ADJ, London WB, et al: The International Neuroblastoma Risk Group (INRG) classification system: An INRG task force report. *J Clin Oncol* 27:289-297, 2009
4. Cheung NKV, Dyer MA: Neuroblastoma: Developmental biology, cancer genomics, and immunotherapy. *Nat Rev Cancer* 13:397-411, 2013
5. Whittle SB, Smith V, Doherty E, et al: Overview and recent advances in the treatment of neuroblastoma. *Expert Rev Anticancer Ther* 17:369-386, 2017
6. Sokol E, Desai AV: The evolution of risk classification for neuroblastoma. *Children (Basel)* 6:27, 2019
7. London WB, Castleberry RP, Matthay KK, et al: Evidence for an age cutoff greater than 365 days for neuroblastoma risk group stratification in the children's oncology group. *J Clin Oncol* 23:6459-6465, 2005
8. Valteau-Couanet D, Schleiermacher G, Sarnacki S, et al: High-risk neuroblastoma treatment strategy: The experience of the SIOPEN group [in French]. *Bull Cancer* 105:918-924, 2018
9. Morgenstern DA, Pötschger U, Moreno L, et al: Risk stratification of high-risk metastatic neuroblastoma: A report from the HR-NBL-1/SIOPEN study. *Pediatr Blood Cancer* 65:e27363, 2018
10. Matthay KK, Reynolds CP, Seeger RC, et al: Long-term results for children with high-risk neuroblastoma treated on a randomized trial of myeloablative therapy followed by 13-cis-retinoic acid: A children's oncology group study. *J Clin Oncol* 27:1007-1013, 2009
11. Ladenstein R, Pötschger U, Pearson ADJ, et al: Busulfan and melphalan versus carboplatin, etoposide, and melphalan as high-dose chemotherapy for high-risk neuroblastoma (HR-NBL1/SIOPEN): An international, randomised, multi-arm, open-label, phase 3 trial. *Lancet Oncol* 18:500-514, 2017
12. London WB, Bagatell R, Weigel BJ, et al: Historical time to disease progression and progression-free survival in patients with recurrent/refractory neuroblastoma treated in the modern era on children's oncology group early-phase trials. *Cancer* 123:4914-4923, 2017
13. Basta NO, Halliday GC, Makin G, et al: Factors associated with recurrence and survival length following relapse in patients with neuroblastoma. *Br J Cancer* 115:1048-1057, 2016
14. Skipper HE, Schabel FM, Wilcox WS: Experimental evaluation of potential anticancer agents XIII. On the criteria and kinetics associated with "curability" of experimental leukemia. *Cancer Chemother Rep* 35:1-111, 1964
15. Altrock PM, Liu LL, Michor F: The mathematics of cancer: Integrating quantitative models. *Nat Rev Cancer* 15:730-745, 2015
16. Benzekry S, Pasquier E, Barbolosi D, et al: Metronomic reloaded: Theoretical models bringing chemotherapy into the era of precision medicine. *Semin Cancer Biol* 35:53-61, 2015
17. Barbolosi D, Ciccolini J, Lacarelle B, et al: Computational oncology—Mathematical modelling of drug regimens for precision medicine. *Nat Rev Clin Oncol* 13:242-254, 2016
18. Benzekry S: Artificial intelligence and mechanistic modeling for clinical decision making in oncology. *Clin Pharmacol Ther*, 108:471, 2020
19. Clancy CE, An G, Cannon WR, et al: Multiscale modeling in the clinic: Drug design and development. *Ann Biomed Eng* 44:2591-2610, 2016

20. Topol EJ: High-performance medicine: The convergence of human and artificial intelligence. *Nat Med* 25:44, 2019
21. Bilous M, Serdjebi C, Boyer A, et al: Quantitative mathematical modeling of clinical brain metastasis dynamics in non-small cell lung cancer. *Sci Rep* 9:13018, 2019
22. Nicolò C, Périer C, Prague M, et al: Machine learning and mechanistic modeling for prediction of metastatic relapse in early-stage breast cancer. *JCO Clin Cancer Inform* 4:259-274, 2020
- 22a. Mechanistic modeling of metastasis for prediction of overall survival in neuroblastoma. [https://github.com/benzekry/mechanistic\\_predict\\_neuroblastoma](https://github.com/benzekry/mechanistic_predict_neuroblastoma)
23. Shulkin BL, Shapiro B: Current concepts on the diagnostic use of MIBG in children. *J Nucl Med* 39:679-688, 1998
24. Matthay KK, Shulkin B, Ladenstein R, et al: Criteria for evaluation of disease extent by 123I-metaiodobenzylguanidine scans in neuroblastoma: A report for the International Neuroblastoma Risk Group (INRG) Task Force. *Br J Cancer* 102:1319-1326, 2010
25. Bleeker G, Tytgat GAM, Adam JA, et al: 123I-MIBG scintigraphy and 18F-FDG-PET imaging for diagnosing neuroblastoma. *Cochrane Database Syst Rev* 9:CD009263, 2015
26. Ladenstein R, Lambert B, Pötschger U, et al: Validation of the mIBG skeletal SIOPEN scoring method in two independent high-risk neuroblastoma populations: The SIOPEN/HR-NBL1 and COG-A3973 trials. *Eur J Nucl Med Mol Imaging* 45:292-305, 2018
27. Benzekry S, Tracz A, Mastri M, et al: Modeling spontaneous metastasis following surgery: An in vivo-in silico approach. *Cancer Res* 76:535-547, 2016
28. Spratt JS, Meyer JS, Spratt JA: Rates of growth of human solid neoplasms: Part I. *J Surg Oncol* 60:137-146, 1995
29. C Davidson-Pilon, J Kalderstam, N Jacobson, et al: CamDavidsonPilon/lifelines: v0.25.4. Zenodo, 2020, <https://zenodo.org/record/4002777#X1j6-GczY8M>
30. Harrell FE, Lee KL, Mark DB: Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 15:361-387, 1996
31. Pang QM, Li K, Ma LJ, et al: Clinical research on neuroblastoma based on serum lactate dehydrogenase. *J Biol Regul Homeost Agents* 29:131-134, 2015
32. Dornenburg C, Fischer M, Barth TFE, et al: LDHA in neuroblastoma is associated with poor outcome and its depletion decreases neuroblastoma growth independent of aerobic glycolysis. *Clin Cancer Res* 24:5772-5783, 2018
33. 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
34. Haeno H, Gonen M, Davis MB, et al: Computational modeling of pancreatic cancer reveals kinetics of metastasis suggesting optimum treatment strategies. *Cell* 148:362-375, 2012
35. Reiter JG, Makohon-Moore AP, Gerold JM, et al: Minimal functional driver gene heterogeneity among untreated metastases. *Science* 361:1033-1037, 2018
36. Newton PK, Mason J, Bethel K, et al: Spreaders and sponges define metastasis in lung cancer: A Markov chain Monte Carlo mathematical model. *Cancer Res* 73:2760-2769, 2013
37. 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
38. Gallik KL, Treffy RW, Nacke LM, et al: Neural crest and cancer: Divergent travelers on similar paths. *Mech Dev* 148:89-99, 2017
39. Borriello L, Seeger RC, Asgharzadeh S, et al: More than the genes, the tumor microenvironment in neuroblastoma. *Cancer Lett* 380:304-314, 2016
40. Vanichapol T, Chutipongtanate S, Anurathapan U, et al: Immune escape mechanisms and future prospects for immunotherapy in neuroblastoma. *Biomed Res Int* 2018:1812535, 2018
41. Jubierre L, Jiménez C, Rovira E, et al: Targeting of epigenetic regulators in neuroblastoma. *Exp Mol Med* 50:51, 2018
42. Benzekry S, Lamont C, Barbolosi D, et al: Mathematical modeling of tumor-tumor distant interactions supports a systemic control of tumor growth. *Cancer Res* 77:5183-5193, 2017
43. Zhang L, Lv C, Jin Y, et al: Deep learning-based multi-omics data integration reveals two prognostic subtypes in high-risk neuroblastoma. *Front Genet* 9:477, 2018
44. Depuydt P, Koster J, Boeva V, et al: Meta-mining of copy number profiles of high-risk neuroblastoma tumors. *Sci Data* 5:180240, 2018
45. Salazar BM, Balczewski EA, Ung CY, et al: Neuroblastoma, a paradigm for big data science in pediatric oncology. *Int J Mol Sci* 18:37, 2016

