

# **True Mitotic Count Prediction in Gastrointestinal Stromal Tumors (GIST): Bayesian Network Model and PROMETheus (PReOperative Mitosis Estimator Tool) App development.**

Salvatore Lorenzo Renne, Manuela Cammelli, Ilaria Santori, Marta Tassan-Mangina, Laura Samà, Laura Ruspi, Federico Sicoli, Piergiuseppe Colombo, Luigi Maria Terracciano, Vittorio Quagliuolo, Ferdinando Carlo Maria Cananzi

Submitted to: Journal of Medical Internet Research  
on: June 16, 2023

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

## Table of Contents

---

<b>Original Manuscript.....</b>	<b>5</b>
<b>Supplementary Files.....</b>	<b>26</b>
Figures .....	27
Figure 1.....	28
Figure 2.....	29
Figure 3.....	30
Figure 4.....	31
Figure 5.....	32
Figure 6.....	33
Multimedia Appendixes .....	34
Multimedia Appendix 0.....	35
Multimedia Appendix 0.....	35
Multimedia Appendix 1.....	35
Multimedia Appendix 2.....	35
Multimedia Appendix 3.....	35
Multimedia Appendix 4.....	35
Multimedia Appendix 5.....	35
Multimedia Appendix 6.....	35
Multimedia Appendix 7.....	35
Multimedia Appendix 8.....	35

# True Mitotic Count Prediction in Gastrointestinal Stromal Tumors (GIST): Bayesian Network Model and PROMETheus (PReOperative Mitosis Estimator Tool) App development.

Salvatore Lorenzo Renne<sup>1, 2</sup> MD; Manuela Cammelli<sup>3\*</sup> MD; Ilaria Santori<sup>3</sup> MD; Marta Tassan-Mangina<sup>3</sup> MD; Laura Samà<sup>4</sup> MD; Laura Ruspi<sup>4</sup> MD; Federico Sicoli<sup>4</sup> MD; Piergiuseppe Colombo<sup>3</sup> MD; Luigi Maria Terracciano<sup>3</sup> MD; Vittorio Quagliuolo<sup>4</sup> MD; Ferdinando Carlo Maria Cananzi<sup>4, 5\*</sup> MD

<sup>1</sup>Department of Biomedical Sciences, Humanitas University Pieve Emanuele, Milan IT

<sup>2</sup>Pathology Department, IRCCS Humanitas Research Hospital Rozzano, Milan IT

<sup>3</sup>Pathology Department, IRCCS Humanitas Research Hospital via Manzoni 56, 20089 Rozzano, Milan IT

<sup>4</sup>Sarcoma, Melanoma and Rare Tumors Surgery Unit, IRCCS Humanitas Research Hospital via Manzoni 56, 20089 Rozzano, Milan IT

<sup>5</sup>Department of Biomedical Sciences, Humanitas University Via Rita Levi Montalcini 4, 20072 Pieve Emanuele, Milan IT

\* these authors contributed equally

## Corresponding Author:

Salvatore Lorenzo Renne MD

Department of Biomedical Sciences, Humanitas University

Via Rita Levi Montalcini 4

Pieve Emanuele, Milan

IT

## Abstract

**Background:** Gastrointestinal Stromal Tumors (GISTs) present a complex clinical landscape, where precise preoperative risk assessment plays a pivotal role in guiding therapeutic decisions. Conventional methods for evaluating mitotic count, such as biopsy-based assessments, encounter challenges stemming from tumor heterogeneity and sampling biases, thereby underscoring the urgent need for innovative approaches to enhance prognostic accuracy.

**Objective:** The primary objective of this study was to develop a robust and reliable computational tool, PROMETheus, aimed at refining patient stratification through precise estimation of mitotic count in GISTs.

**Methods:** Leveraging advanced Bayesian Network methodologies, we constructed a Directed Acyclic Graph (DAG) integrating pertinent clinico-pathological variables essential for accurate mitotic count prediction on the surgical specimen. Key parameters identified and incorporated into the model encompassed tumor size, location, mitotic count from biopsy specimens, surface area evaluated during biopsy, and tumor response to therapy, when applicable. Rigorous testing procedures, including prior predictive simulations, validation utilizing synthetic datasets were employed. Finally, the model was trained on a comprehensive cohort of real-world GIST cases (n=80), drawn from the repository of IRCCS Humanitas Research Hospital, totaling 160 cases, were conducted.

**Results:** Our computational model exhibited excellent diagnostic performance on synthetic data, different model architecture were selected based on lower deviance and robust out-of-sample predictive capabilities. Posterior predictive checks (retrodiction) further corroborated the model's accuracy. Subsequently, the development of PROMETheus, an intuitive application dynamically computing predicted mitotic count and risk assessment on surgical specimens based on tumor-specific attributes, including size, location, surface area, and biopsy-derived mitotic count, using posterior probabilities derived from the model, was successfully achieved.

**Conclusions:** The deployment of PROMETheus might herald a significant advancement in preoperative risk stratification for GISTs, offering clinicians a precise and reliable means to anticipate mitotic counts on surgical specimens and a solid base to stratify patients for clinical studies. By facilitating tailored therapeutic strategies, this innovative tool is poised to revolutionize clinical decision-making paradigms, ultimately translating into improved patient outcomes and enhanced prognostic precision in the management of GISTs.

(JMIR Preprints 16/06/2023:50023)

DOI: <https://doi.org/10.2196/preprints.50023>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

## Original Manuscript

# True Mitotic Count Prediction in Gastrointestinal Stromal Tumors (GIST): Bayesian Network Model and PROMETheus (PReOperative Mitosis Estimator Tool) App developement.

Salvatore Lorenzo Renne<sup>1</sup>

Manuela Cammelli<sup>2</sup>

Ilaria Santori

Marta Tassan-Mangina

Laura Ruspi

Federico Sicoli

Laura Samà

Piergiuseppe Colombo

Luigi Maria Terracciano

Vittorio Quagliuolo

Ferdinando Carlo Maria Cananzi<sup>3</sup>

2024-05-13

---

<sup>1</sup> Corresponding author

<sup>2</sup> Co-first author

<sup>3</sup> Co-corresponding author

**Background:** Gastrointestinal Stromal Tumors (GISTs) present a complex clinical landscape, where precise preoperative risk assessment plays a pivotal role in guiding therapeutic decisions. Conventional methods for evaluating mitotic count, such as biopsy-based assessments, encounter challenges stemming from tumor heterogeneity and sampling biases, thereby underscoring the urgent need for innovative approaches to enhance prognostic accuracy. **Objective:** The primary objective of this study was to develop a robust and reliable computational tool, PROMETheus, aimed at refining patient stratification through precise estimation of mitotic count in GISTs. **Methods:** Leveraging advanced Bayesian Network methodologies, we constructed a Directed Acyclic Graph (DAG) integrating pertinent clinico-pathological variables essential for accurate mitotic count prediction on the surgical specimen. Key parameters identified and incorporated into the model encompassed tumor size, location, mitotic count from biopsy specimens, surface area evaluated during biopsy, and tumor response to therapy, when applicable. Rigorous testing procedures, including prior predictive simulations, validation utilizing synthetic datasets were employed. Finally, the model was trained on a comprehensive cohort of real-world GIST cases (n=80), drawn from the repository of IRCCS Humanitas Research Hospital, totaling 160 cases, were conducted. **Results:** Our computational model exhibited excellent diagnostic performance on synthetic data, different model architecture were selected based on lower deviance and robust out-of-sample predictive capabilities. Posterior predictive checks (retrodiction) further corroborated the model's accuracy. Subsequently, the development of PROMETheus, an intuitive application dynamically computing predicted mitotic count and risk assessment on surgical specimens based on tumor-specific attributes, including size, location, surface area, and biopsy-derived mitotic count, using posterior probabilities derived from the model, was successfully achieved. **Conclusions:** The deployment of PROMETheus represents a potential advancement in preoperative risk stratification for GISTs, offering clinicians a precise and reliable means to anticipate mitotic counts on surgical specimens and a solid base to stratify patients for clinical studies. By facilitating tailored therapeutic strategies, this innovative tool is poised to revolutionize clinical decision-making paradigms, ultimately translating into improved patient outcomes and enhanced prognostic precision in the management of GISTs.

## Introduction

Gastrointestinal Stromal Tumor (GIST) is the most common sarcoma type [1]. The majority harbors activating mutation in *KIT* or *PDGFRA* [2–5]. Even if these mutations represent early events in carcinogenesis – being shared by clinically irrelevant and very aggressive GIST [6,7] – they are the molecular basis for the very active tyrosine-kinase-inhibitor (TKI) therapy [8–10]. TKI revolutionized GIST treatment and they have been used in all the disease stages since their introduction [11–14]. Notably, they can be administered as neo-adjuvant treatment for patients with high-risk disease or to reduce the extent of the surgery in GIST in peculiar location (i.e. rectum and duodenum) [14–17]. Even if neo-adjuvant therapy with Imatinib is beneficial in patients with high risk disease [18], and in patients that will have not an R0 surgery or patients that can have a less mutilating, function-sparing surgery if there is a volumetric reduction [17], TKI has some issues: on the one hand it is likely to impair a correct – post operative – risk assessment: indeed if the tumor responds to therapy it will be not possible to correctly assess the risk [5]. Several risk assessments in GIST have been developed and they identify size, site and mitotic count as important features [19–21]. Indeed, mitotic count on the surgical specimen after TKI therapy can be greatly modified – especially in the case of tumor response – therefore in these patients the subsequent management will be guided by a risk assessment computed with the mitotic count from the biopsy, and this – we will see – can lead to mistreatment.

On the other hand, during initial patient management, identification of the high-risk patients can also fail: a very small amount of tissue on biopsy is required to make the diagnosis of GIST, since very good immunohistochemical markers (rather specific and sensitive) exist [22–25]. Therefore, whereas size and site can be accurately assessed by imaging, mitotic count on biopsy can be burdened by several limitations: some purely biological – such as tumor heterogeneity –, others more physical, i.e. the size of the specimen available

for counting (a classic example of *sampling bias*) [26–30]. Thus, it may happen to incorrectly classify the risk of a GIST preoperatively, and this might lead to *surprises* after the mitotic count on the surgical specimen is performed, often due to underestimation of the mitotic count [31–33].

These limitations underscore the critical need for innovative approaches to refine preoperative risk stratification in GIST, aiming to mitigate the risks of misclassification and subsequent therapeutic mismanagement. The discrepancies between preoperative risk assessments and postoperative findings underscore the imperative for precision tools that can dynamically estimate mitotic count on surgical specimens, enhancing the accuracy of patient stratification and treatment planning.

In line with this imperative, we aim to develop an advanced computational tool, termed PROMETheus, designed to predict mitotic count on surgical specimens. By leveraging state-of-the-art Bayesian modeling techniques and integrating comprehensive clinico-pathological variables, PROMETheus seeks to address the limitations of current risk assessment methodologies, offering clinicians a reliable means to anticipate postoperative mitotic counts and refine preoperative treatment strategies effectively.

## Materials and methods

### Modeling strategy

**Bayesian network and workflow.** As modeling strategy we used Bayesian network with the aim of predicting the mitotic count on the surgical specimen. Of note, we use the term *Bayesian network* to indicate the graphical representation of the model and the collection of *functions* necessary to use it for statistical learning. This looser definition is often used in practice, however it's broader than the one indented by the term's inventor, Judea Pearl. In its *Causality* (2009) he basically identifies the directed acyclic graphs (DAGs, see main text) with the term Bayesian networks (BN): “Directed graphs, especially DAGs, have been used to represent causal or temporal relationships [...] and came to be known as Bayesian networks, a term coined [...] to emphasize three aspects: (1) the subjective nature of the input information; (2) the reliance on Bayes's conditioning as the basis for updating information; and (3) the distinction between causal and evidential modes of reasoning ...” – see p. 14 of the cited reference [34]. In the present paper, the meaning of BN is closer to the one of structural causal models (SCMs) [35]. Briefly, we designed a SCM of the variables, created a mock data-set, wrote a probabilistic program, validated it on the data simulation, fit the model to the data, and compared multiple models with different structures; these procedures are often collectively referred as the *Bayesian workflow* [36–39]. **Causal modeling.** The graphical representation of the Bayesian network was done with Directed Acyclic Graphs (DAG): the *variables* were represented by nodes and the *conditional dependencies* through directed edges. Based on the graphical representation we built a Structural Causal Model (SCM), that we used for data simulation and as the model for the fits. **Probabilistic programming.** We wrote and fit the models using R version 4.1.2 and Stan version 2.21.0 [40–43]. Stan is a probabilistic programming language that runs a No U-Turn sampler, an extension to Hamiltonian Monte Carlo (HMC)



sampling, which is itself a form of Markov Chain Monte Carlo (MCMC) [44]. To promote regularization and reduce overfitting, we used multilevel-hierarchical modeling strategy [39,45,46]. Distributions (likelihood and priors) were chosen with maximum entropy criteria [39]. To understand priors' implications, we run prior predictive simulations [37,38,46]. To minimize divergent transitions, we *reparametrized* the models with a non-centered equivalent form when appropriated [39,47]. To ensure a good representation of the sample space, we visually inspected the chains with traceplots and trunkplots [48]. We then monitored the chains with post-modeling diagnostics such as the number of effective samples and the Gelman-Rubin convergence diagnostic  $\hat{R}$  [45,48]. Of note,  $\hat{R}$  as defined in the cited references is different from the classic definition by Gelman and Rubin (1992). All models' fits were plotted against the fitted data to ensure a good representation of the outcome space (posterior predictive check) [39,49,50]. Compatibility intervals (CIs) were calculated as Highest Posterior Density Interval (HDPI) [51]. **Model selection.** For the aim of pure prediction – as in our case – the best model can be selected based on information theory estimating model performance through Widely Applicable Information Criteria (WAIC, a generalization of Akaike information criteria) and Pareto Smooth Important Sampling Leave-One-Out Cross-Validation Criteria (PSIS-LOO-CV) [52–54]. We checked that the two statistics gave the same results to *trust* their results [39,55,56]. We selected the model with the lowest deviance in out of sample performance. **Forecasting.** The posterior probability density of the coefficients was then used to create an application that, given the chosen variables, computed the posterior probability distribution in the outcome space (i.e. the mitotic count on the specimen). Moreover, we programmed the app to calculate the risk class from this computed posterior distribution of the mitotic count.

## Study population

**Dataset.** The cases came from a prospectively maintained database including all the patients who underwent surgery for primary sarcoma in IRCCS Humanitas Research Hospital (Rozzano (MI), Italy). This database comprises extended clinical and pathological information and contained 233 GISTs operated from January 2000 to March 2022. **Inclusion Criteria.** We included patients with pre-operative diagnostic biopsy, that underwent surgical resection, with informed consent to research and available histologic material of the biopsy and the surgical specimen. **Exclusion criteria.** Histology was reviewed by a sarcoma pathologist (SLR) and cases with diagnosis other than GIST were excluded.

## Pathology

**Microscope calibration.** We calibrated the microscope with a stage micrometer slide and calculated the number of high power fields (HPFs) needed to reach the size of  $5 \text{ mm}^2$  and – in line with published guidelines – the number of HPFs to evaluate was 23.5 [30,57–59]. **Mitosis.** We defined a mitosis as basophilic, dark, hairy material representing the chromosomes. A mitosis was counted when the *chromosomes* were either clotted (as in the beginning of metaphase), in a plane (as in metaphase and anaphase), or in separate clots (as in telophase), as previously described[60]. **Biopsy measurement.** We measured the surface available under the microscope counting HPFs filled by neoplastic

specimen up to 5  $\text{mm}^2$ ; for very small biopsies we approximated the surface as fraction of a field of view.

**Tumor response.** Some of the cases of the series underwent preoperative therapy. To use these cases without polluting the estimate for mitotic count coefficient, we also recorded the *response to treatment*; this had a different meaning from a classical pathological response and was defined as follows: if a mitotically active area was identified on the surgical specimen (regardless of the size) and the mitotic count in this area was equal to or more than the biopsy count the tumor was classified as *no response*. Therefore this mitotic count on the biopsy was used in the model as if the case did not undergo preoperative therapy. Conversely, if the count on the surgical specimen was less than the count on the biopsy, the tumor was classified as *response* and the count was used to estimate a different coefficient that we did not use for prediction (see model description in the result section and custom code for greater detail).

## Ethical Considerations

Patients signed an institutional written informed consent to research. According to European regulation (UE2016/679) and Italian Privacy Code (D.Lgs.101/2018) retrospective observational monocentric study do not require specific informed consent. Data were de-identified prior to analysis.

## Results

### Causal modeling

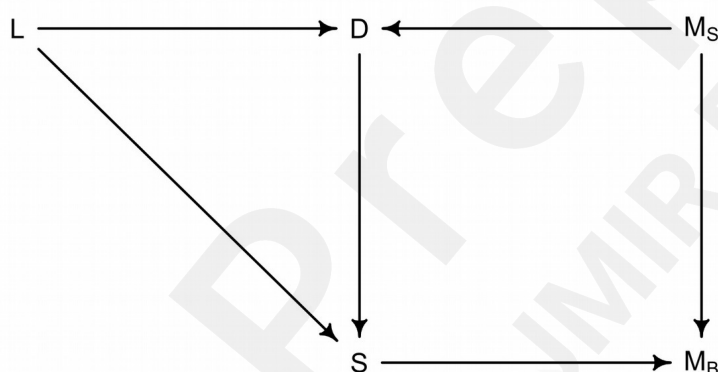


Figure 1: To determine the model covariates, we utilized a Directed Acyclic Graph (DAG). In this causal framework, several factors influence the mitotic count on the surgical specimen. Firstly, the dimension of the tumor impacts both the mitotic count on the biopsy and the surgical specimen. Larger tumors tend to exhibit higher mitotic activity. Secondly, the location of the tumor plays a crucial role in its growth pattern. For instance, gastric neoplasms often have more space to expand, leading to the development of symptoms with larger masses. Additionally, tumor location influences the accessibility of the biopsy site, as some sites are inherently more accessible than others. Lastly, the amount of surface area available on the biopsy directly impacts the accuracy of the mitotic count estimation. A larger surface area allows for more representative sampling of the tumor. This simplified causal model elucidates the relationship between various factors affecting GIST mitotic count. Please refer to the main text and supplementary materials for further details. In the model notation,  $D$  represents the dimension of the tumor,  $L$  denotes the location,  $M_B$  signifies the mitotic count on the biopsy,  $M_S$  indicates the mitotic count on the surgical specimen, and  $S$  represents the surface area of the biopsy.

We designed a causal model to identify the covariates to be include in the model for estimating the mitotic count on the surgical specimen. We assumed that the true mitotic count is the one counted on the surgical specimen ( $M_s$ ). The more tumor cell replicates, the bigger the tumor is ( $D$ ). We also identified the anatomical location ( $L$ ) as a cause of tumor dimension ( $D$ ) in the sense that in certain locations the symptoms would appear earlier thus influencing the measured size at diagnosis. Moreover, location directly *causes* the total amount of tissue available for evaluation at biopsy (i.e. the measured biopsy surface,  $S$ ): some sites are more difficult to reach than others. Lastly, mitotic count on the biopsy ( $M_b$ ) reflects the mitotic count on the surgical specimen ( $M_s$ ) and due to tumor heterogeneity and sampling bias the measurement on the biopsy also depends on the surface examined ( $S$ ) (Figure 1 and also figure 8 for an extended version of the DAG).

## Probabilistic modeling

Our inferential target was the mitotic count on the surgical specimen ( $M_s$ ); as the name suggests it is a count variable and we therefore chose a Poisson distribution to model it (Equation (1)). Poisson distributions have just one parameter,  $\lambda$ . It is the expected value and the expected variance of the count variable and it is the parameter used for the generalized linear model. It needs to be positive, and a common link function is to exponentiate the model. Given each patient  $i$ , we estimated a coefficient for the tumor dimension ( $\beta$ ) and for surface of the biopsy ( $\gamma$ ) for each location ( $L$ ) using a multilevel-hierarchical model for both of them ( $\beta_{[L]}D_i$  and  $\gamma_{[L]}S_i$  respectively); the parameters  $\delta$  and  $\epsilon$  were alternatively switched on and off by the presence of response to therapy  $R$ , as defined in the methods section; and finally we set an intercept  $\alpha$  (Equation (2)). To justify prior choice we used prior predictive simulation.

$$\begin{aligned}
 M_{Si} & \sim \text{Poisson}(\lambda_i) (1) \\
 \log(\lambda_i) & \sim \alpha + \beta_{[L]}D_i + \gamma_{[L]}S_i + \delta M_{Bi}(1 - R_i) + \epsilon M_{Bi}R_i (2) \\
 \alpha & \sim N(-1, 0.2) (3) \\
 \beta_{[L]} & \sim N(\hat{\beta}, \sigma_\beta) (4) \\
 \hat{\beta} & \sim N(-1, 0.2) (5) \\
 \sigma_\beta & \sim \text{Exponential}(2) (6) \\
 \gamma_{[L]} & \sim N(\hat{\gamma}, \sigma_\gamma) (7) \\
 \hat{\gamma} & \sim N(-1, 0.2) (8) \\
 \sigma_\gamma & \sim \text{Exponential}(2) (9) \\
 \delta & \sim N(-1, 0.2) (10) \\
 \epsilon & \sim N(-1, 0.2) (11)
 \end{aligned}$$

## Prior predictive simulation

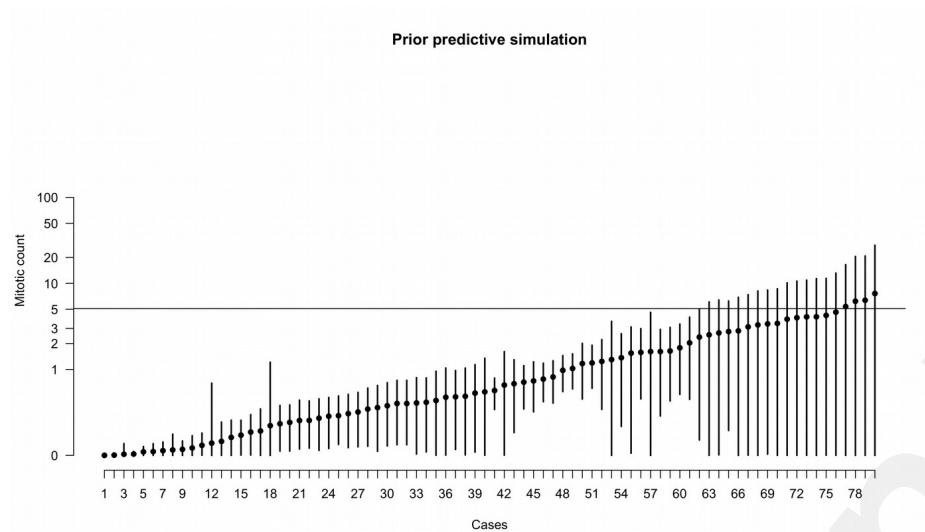


Figure 2: Prior predictive simulation of the  $\lambda$  parameter, i.e., what the model anticipates before encountering the data. This plot displays 80 simulated cases derived from the priors. The majority of expected values indicate a very low mitotic count, aligning with real-world expectations. However, the model is not startled by higher mitotic counts, even though it anticipates encountering them in a minority of cases without prior training.

The majority of GISTs are clinically irrelevant, with a mitotic count less than  $5/5 \text{ mm}^2$ . Few of them can have an higher mitotic count, even to a greater order of magnitude, however it is biologically implausible to expect many cases with mitotic count greater than 50. Using this field specific knowledge, we chose normal distributions for model coefficients (Equations (3), (5), (4), (7), (8), (10), (11)) and exponential distributions for scalar coefficients (Equations (6), (9)). Through serial simulations we narrowed the numerical values. For a graphical representation of part of the coefficients see figure 9. The results of the prior predictive simulation show that the most of the simulated  $\lambda$  have indeed values in keeping with the field specific knowledge (Figure 2).

## Fitting the mock data-set

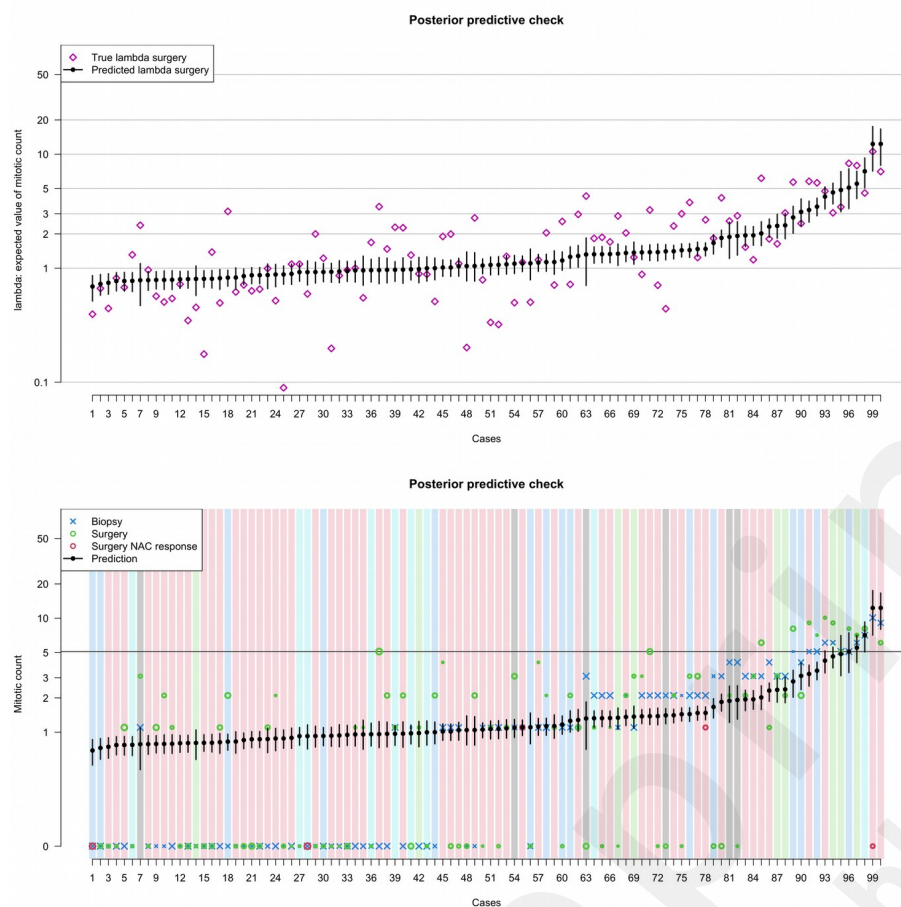


Figure 3: Posterior predictive simulation. The upper panel shows the imputed  $\lambda$  parameter for each sample against the ground-truth; consider the logarithmic scale on the y-axis. The lower panel shows the same imputed  $\lambda$  parameter against the fitted data. On average the mitotic count either on the surgical specimen or on the biopsy is lower than the expected value, this is due to the asymmetry of the Poisson distribution. The size of the X is proportional to the biopsy surface, the size of the empty dots is proportional to the tumor size; the color of the background corresponds to the tumor site.

To test the model performance, we fitted it on a simulated data-set of 100 cases. The custom code to produce the mock data-set is available on the cited repository. The fit with a centered parameterization resulted in 2% of divergent transitions, we therefore rewrote the model in a non-centered form. Fit's diagnostics were satisfactory:  $\hat{R}=1.00$  was obtained for all the parameters, the energy from the Hamiltonian had a Gaussian outlook and the trunkplots of the log-probability showed a satisfactory convergence of the chains (Figure 10); all the parameters had a satisfactory number of effective samples and a good outlook of the trunkplots (Figure 11); the posterior probability density for each coefficient is depicted in figure 12. To check the fitness of the model, we compared the inferred  $\lambda$  by the model to the true  $\lambda$  used for data simulation. This procedure revealed that the model regularized the values within each modeled site and was able to recover a value closer to  $\lambda$  even if provided with lower values (Figures 3).

## Fitting the real data-set

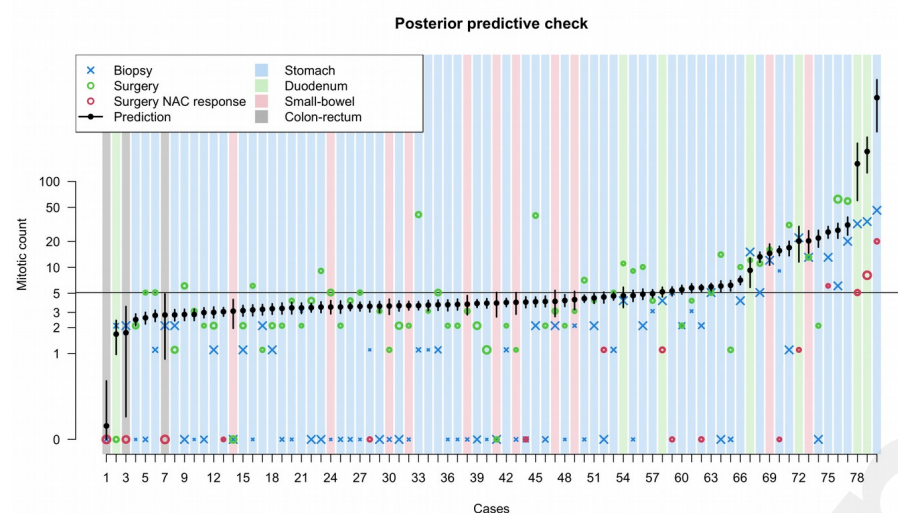


Figure 4: Posterior predictive simulation on the true dataset. The size of the X is proportional to the biopsy surface, the size of the empty dots is proportional to the tumor size; the color of the background corresponds to the tumor site.

We hence fitted the model to the real data. We were able to fit it with the centered version of the model. Similarly to the simulated dataset, fit's diagnostics were satisfactory:  $\hat{R}=1.00$  was obtained for all the parameters, the energy from the Hamiltonian had a Gaussian outlook and the trunkplots of the log-probability showed a satisfactory convergence of the chains (Figure 13); moreover, all the parameters had a satisfactory number of effective samples and a good outlook of the trunkplots (Figure 14); the posterior probability density for model coefficient is depicted in figure 15 and table 1. The site ( $L$ ) was an index variable with values from 1 to 4, representing coefficients for colon-rectum, duodenum, small-intestine and stomach, respectively. Of note the model posterior distribution for the biopsy count parameter ( $\delta$ ) is 1, consistently with our expectations. As in the data simulation, in order to see what the model *learnt*, we moved to form the parameter space to the outcome space and simulated from the whole posterior distribution a  $\lambda$  parameter – that tells us the expected value of the Poisson distribution – for each case and plotted it against the data fitted. This posterior predictive simulation is depicted in figure 4.

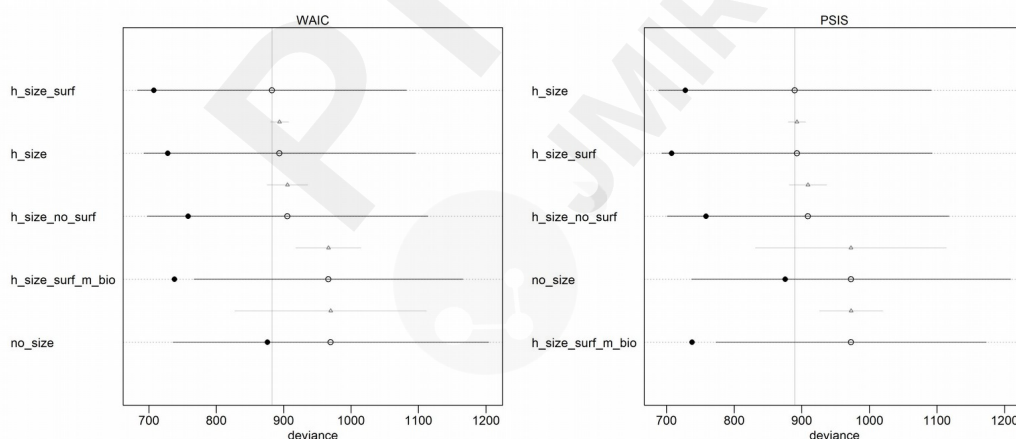
Table 1: Coefficients

	mean	SD	5.5%	94.5%	N eff.	$\hat{R}$
$\beta_{L[1]}$	3.81	1.91	1.40	7.10	4396	1
$\beta_{L[2]}$	0.12	0.15	-0.12	0.36	8407	1
$\beta_{L[3]}$	0.30	0.16	0.03	0.55	5850	1
$\beta_{L[4]}$	0.40	0.04	0.35	0.46	9032	1
$\gamma_{L[1]}$	0.38	0.44	-0.25	1.13	6905	1
$\gamma_{L[2]}$	-0.09	0.15	-0.33	0.15	8769	1
$\gamma_{L[3]}$	-1.22	0.34	-1.77	-0.70	5884	1

	mean	SD	5.5%	94.5%	N eff.	$\hat{R}$
$\gamma_{L[4]}$	-0.17	0.06	-0.26	-0.08	9371	1
$\alpha$	1.75	0.05	1.67	1.83	10178	1
$\dot{\gamma}$	-0.85	0.20	-1.17	-0.53	8664	1
$\epsilon$	0.16	0.05	0.08	0.23	7978	1
$\delta$	1.01	0.08	0.89	1.13	8031	1
$\dot{\beta}$	-0.90	0.20	-1.22	-0.57	9447	1
$\sigma_{\gamma}$	0.89	0.36	0.44	1.55	6856	1
$\sigma_{\beta}$	2.07	0.74	1.10	3.40	5274	1

## Model selection

In general, adding parameters in multilevel modeling reduces overfitting and improves out-of-sample performance, whereas adding parameters without a hierarchical structure can reduce deviance within sample, but results in a lower out-of-sample performance (i.e. increases overfitting). We therefore computed the deviance using WAIC and PSIS-LOO-CVC on our model (that had a multilevel – hierarchical – structure for the tumor size  $\beta_L$  and biopsy surface  $\gamma_L$  parameters) and for alternative models having also the mitotic count on the biopsy with a hierarchical structure (see Equation (12)), only tumor size with a hierarchical structure (see Equation (13)), only tumor size and without accounting for biopsy surface (see Equation (14)), and no hierarchical structure without accounting for tumor size (see Equation (15)). The model with hierarchical structure for tumor size and biopsy surface parameters (Equation (2)) had the lowest out-of-sample deviance (Figure 5). We therefore chose the model with hierarchical structure for the tumor size and biopsy surface parameters for the App development.

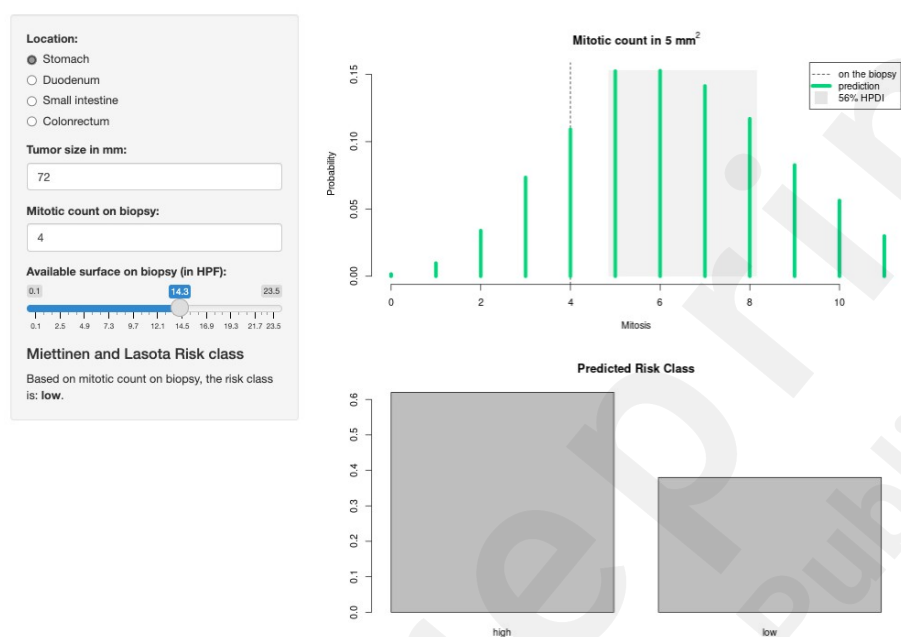


**Figure 5: Model Selection.** The vertical line indicates the mean deviance of the reference model (the one with the lowest out of sample deviance). The filled dots are values within the sample. The empty dots represent mean the out of sample deviance with the bar indicating the 89% CI; the triangle is the contrast between the model and the reference model.



## Application

Using the posterior probabilities from the model, we developed PROMETheus (PReOperative Mitosis Estimator Tool), a web-based application freely available at <https://slrenne.shinyapps.io/PROMETheus/>. The user interface has an input panel to insert the tumor location, the tumor size, the mitotic count on the biopsy, and the available surface on the biopsy. Then the app dynamically computes the risk class according to Miettinen and Lasota. Moreover, using the inputted data and the full posterior distribution, the App computes the expected mitotic count on the surgical specimen (indicating also most probable rendered counts) and the predicted risk class for the new posterior distribution of  $M_s$  provided (Figure 6).



*Figure 6: The PROMETheus (PReOperative Mitosis Estimator Tool) App. The image shows a screenshot for a gastric GIST measuring 72 mm with a biopsy surface of 14.3 HPFs (about 3 mm<sup>2</sup>). With these biological characteristics the risk class is **low**. However the model shows that the mitotic count is likely to be underestimated, with the most probable count on the surgical specimen predicted to be between 5 to 8 mitoses in 5 mm<sup>2</sup>. Given this predicted mitotic count distribution, it is much more probable for the risk class to be high (more than 60%). See text for further explanation.*

## Development of the Preoperative Classification for GIST

To facilitate stage-adapted treatment planning and enhance data comparison across institutions, we developed a preoperative classification system for GIST based on the results of our computational model, PROMETheus. This classification system aims to provide a standardized definition of preoperative classification for GISTs, which can be used to guide therapeutic decisions and improve clinical trial designs (Table 2). The development of this classification included the probabilistic output of PROMETheus risk classification, important surgical parameters (such as site, size, etc.), and resectability.



Table 2: Proposed Preoperative Classification for GIST.<sup>4</sup>

Clinical

Type	GIST's Resectability	Features
0	Resectable	Gastric, mini-tumor (< 2 cm), PROMETheus Low Risk
1	Resectable	Gastric, small tumor (2–10 cm), PROMETheus Low Risk
2	Resectable locally advanced	Large tumor (Gastric: > 10 cm; non-gastric: > 5 cm.) Non-gastric site Elevated risk of intra-operative rupture Need to mutilating or complex multivisceral resection PROMETheus High Risk
3	Resectable metastatic	Resectable synchronous metastatic disease
4	Unresectable	Unresectable non metastatic disease

## Discussion

### Principal Findings

Our study aimed to develop an innovative computational tool, PROMETheus, to accurately predict mitotic count on surgical specimens in Gastrointestinal Stromal Tumors (GISTs), thereby addressing the challenges associated with preoperative risk assessment and treatment planning. The primary objective was to bridge the gap between clinical judgment and computed risk class, empowering clinicians to make informed decisions in complex scenarios. Through the utilization of Bayesian Networks and rigorous covariate selection methodologies, we succeeded in achieving this objective, providing clinicians with a novel approach to preoperative risk stratification in GISTs.

Our findings represent a potential advancement in the field of oncology, particularly in the context of sarcomas, where precise risk assessment is paramount for optimal treatment outcomes. By accurately predicting mitotic count on surgical specimens, PROMETheus offers clinicians a tool to navigate the complexities of GIST management, enabling tailored treatment strategies based on individual patient characteristics. This approach aligns with the evolving paradigm of precision medicine, where treatment decisions are increasingly guided by molecular and pathological insights.

Mitotic count serves as a vital indicator of biological aggressiveness in oncology, playing a significant role in the grading systems of various tumors. With chemotherapy often administered to patients with high-grade tumors, mitotic count becomes a *de facto* predictive biomarker, providing valuable insights into treatment response. In the realm of Gastrointestinal Stromal Tumors (GISTs), alongside size and site, the number of

<sup>4</sup> Adapted from Cananzi et al. incorporating the probabilistic estimate of the Risk Class by our app PROMETheus [26].

mitoses holds utmost importance in current risk classifications [19–21]. However, unlike size and site, which can be easily assessed through imaging or endoscopy in the preoperative setting, accurately determining mitotic count poses challenges. Factors such as tumor heterogeneity and limited specimen size for counting (a classic example of sampling bias) hinder precise estimation during preoperative biopsy.

The introduction of effective therapies like Imatinib, which targets the molecular alterations driving GIST, has led to a clinical tendency to use this treatment preoperatively, even in cases where the risk of disease progression is not high. However, this approach has several drawbacks: Variable Response: not all patients exhibit a decrease in tumor size, which means the extent of the surgery may remain unchanged. Side Effects: Imatinib is not without side effects, highlighting the need for a tool to identify patients at genuine risk of metastasis who would benefit from neoadjuvant treatment. Risk Classification Post-Treatment: for patients treated with neoadjuvant Imatinib, it becomes challenging to perform risk classification on the surgical specimen, leaving uncertainty about the subsequent adjuvant therapy.

The integration of PROMETheus into clinical practice could have far-reaching implications for patient care. Not only does it could enable more precise risk stratification in GISTs, but it also could facilitate stage-adapted treatment planning, ensuring that patients receive the most appropriate interventions based on their individual risk profiles. The preoperative classification system provided in Table 2 offers a common language for patient follow-up and treatment planning. It is a revision of previous work by our group [26], refined based on the insights gained from our current study. This classification could lay the foundation for improved data collection and comparison across institutions, fostering collaboration and advancing research efforts in the field of GIST management.

## Comparison to Prior Work

Importantly, our study diverges from previous approaches that primarily focused on comparing biopsy and surgical specimens to evaluate the reliability of biopsy-based risk assessments. Instead, we recognized the inherent limitations of preoperative biopsy in predicting tumor grade, especially in the context of TKI therapy-induced tumor response. By focusing on predicting mitotic count directly, our approach transcends these limitations, providing clinicians with a more accurate and reliable tool for preoperative risk assessment.

There are many tools for risk stratification in oncological practice, also in the sarcoma field, mainly based on nomograms, that are usually employed to predict overall survival and the risk of metastasis [61]. While these tools have found utility in re-analyzing previous clinical studies and selecting patients for future trials [62], they are not tailored for Gastrointestinal Stromal Tumors (GISTs). Our innovative approach stands out in two crucial ways. First, our Bayesian methodology empowers clinicians by providing the full posterior probability, capturing the inherent uncertainty often overlooked by traditional *frequentist* approaches that focus on central estimates. Second, unlike prognostic prediction tools, our method solely aims to forecast the mitotic count on the surgical specimen, a distinctive objective.

## Strengths and Limitations

Given the widespread use of mitotic count as a grading parameter, our approach holds potential for scalability across diverse clinical settings, such as breast cancer, solitary fibrous tumor, and soft tissue sarcoma [30,63]. The integration of our posterior probability with other tools could enhance their effectiveness. However, our choice of prediction methods was largely constrained by the dataset's size, as machine learning methods like deep learning demand substantial computational resources and larger sample sizes [64], which are often limited in rare diseases. Bayesian networks offer an advantageous alternative with principled variable selection, interpretability, and no minimum sample size requirement, making them a fitting choice for our study.

Despite the promising findings of our study, several limitations warrant consideration.

First, our use of a complex multilevel hierarchical model, meticulously designed according to the Directed Acyclic Graph (DAG), introduces potential challenges in model efficacy and implementation. However, we addressed this complexity by subjecting the model to rigorous testing with simulated data, ensuring its robustness and reliability.

Second, the performance evaluation of PROMETheus is based on data from a relatively limited patient population, which may restrict the generalizability of our findings. However, unlike conventional approaches that rely solely on in-sample performance metrics for model selection, we employed PSIS-LOO-CV (Pareto-smoothed importance sampling leave-one-out cross-validation) to identify the most effective model structure. Nonetheless, further validation on a larger and more diverse population is essential before considering the clinical deployment of PROMETheus, ensuring its efficacy and applicability across different clinical settings and patient demographics.

Third the development and deployment of PROMETheus as a web-based application are at an early stage. Although we have made the tool publicly available as an open-source resource to encourage validation and application by other researchers and clinicians, the current version may lack certain functionalities and user-friendly features (such as inclusion of risk classification other than Miettinen and Lasota). Future work should focus on enhancing the application's interface, usability, and integration with clinical workflows, as well as providing comprehensive user training and support to facilitate its adoption in clinical practice.

## Future Directions

We have developed a cutting-edge App that could revolutionizes the estimation of mitotic count on surgical specimens, not only providing accurate quantification but also addressing the uncertainty that clinicians face when encountering challenging cases. By bridging the gap between clinical judgment and computed risk class based on available parameters, our App empowers clinicians to make informed decisions in complex scenarios. While previous studies have primarily focused on comparing biopsy and surgical specimens to evaluate the reliability of the former as the *gold standard*, we diverge from this approach. Our study

transcends the limitations of preoperative biopsy in predicting tumor grade, shedding light on the underestimation of aggressiveness often associated with this method[65–68]. Embracing the widely accepted practice of causal modeling for covariate selection in epidemiological studies [69], we harness the power of Bayesian Networks to pioneer a novel tool capable of predicting mitotic count—a breakthrough unprecedented in existing literature.

In conclusion, our study represents a significant step forward in the development of precision tools for oncological risk assessment. By accurately predicting mitotic count on surgical specimens in GISTs, PROMETheus empowers clinicians to make informed treatment decisions, ultimately improving patient outcomes and advancing the field of sarcoma management. Moving forward, continued research and validation efforts will be essential to further refine and optimize the utility of PROMETheus in clinical practice, ultimately realizing its full potential in guiding personalized treatment strategies for patients with GISTs and other sarcomas.

## Acknowledgements

We acknowledge the reSeARChOMA group of IRCCS Humanitas Research Hospital.

## Data availability

**Data.** De-identified patient data are available upon motivated request on the public repository Zenodo <https://zenodo.org/communities/humanitasirccs/>. **Code.** Custom code for the analysis is available at [www.github.com/slrenne/PROMETheus](https://www.github.com/slrenne/PROMETheus).

## References

1. Pinieux G de, Karanian M, Le Loarer F, Le Guellec S, Chabaud S, Terrier P, Bouvier C, Batistella M, Neuville A, Robin Y-M, Emile J-F, Moreau A, Larousserie F, Leroux A, Stock N, Lae M, Collin F, Weinbreck N, Aubert S, Mishellany F, Charon-Barra C, Croce S, Doucet L, Quintin-Rouet I, Chateau M-C, Bazille C, Valo I, Chetaille B, Ortonne N, Bouchet A, Rochaix P, Demuret A, Ghnassia J-P, Mescam L, Macagno N, Birtwisle-Peyrottes I, Delfour C, Angot E, Pommepuy I, Ranchere D, Chemin-Airiau C, Jean-Denis M, Fayet Y, Courrèges J-B, Mesli N, Berchoud J, Toulmonde M, Italiano A, Le Cesne A, Penel N, Ducimetiere F, Gouin F, Coindre J-M, Blay J-Y. Nationwide incidence of sarcomas and connective tissue tumors of intermediate malignancy over four years using an expert pathology review network. Pasquali S, editor. PLOS ONE 2021 Feb;16(2):e0246958. PMID:[33630918](#)
2. Hirota S, Isozaki K, Moriyama Y, Hashimoto K, Nishida T, Ishiguro S, Kawano K, Hanada M, Kurata A, Takeda M, Tunio GM, Matsuzawa Y, Kanakura Y, Shinomura Y, Kitamura Y. Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. Science 1998;279(5350):577–580. PMID:[9438854](#)
3. Joensuu H, Hohenberger P, Corless CL. Gastrointestinal stromal tumour. The Lancet Elsevier Ltd; 2013;382(9896):973–983. PMID:[23623056](#)
4. Rubin BP, Heinrich MC. Genotyping and immunohistochemistry of gastrointestinal stromal tumors: An update. Seminars in Diagnostic Pathology Elsevier; 2015;32(5):392–399. PMID:[25766843](#)
5. Mehren M von, Joensuu H. Gastrointestinal stromal tumors. Journal of Clinical Oncology 2018;36(2):136–143. PMID:[29220298](#)
6. Rossi S, Gasparotto D, Toffolatti L, Pastrello C, Gallina G, Marzotto A, Sartor C, Barbareschi M, Cantaloni C, Messerini L, Bearzi I, Arrigoni G, Mazzoleni G, Fletcher JA, Casali PG, Talamini R, Maestro R, Maestra R, Dei Tos AP. Molecular and clinicopathologic characterization of gastrointestinal stromal tumors (GISTs) of small size. The American journal of surgical pathology 2010 Oct;34(10):1480–91. PMID:[20861712](#)
7. Heinrich MC, Patterson J, Beadling C, Wang Y, Debiec-Rychter M, Dewaele B, Corless CL, Duensing A, Raut CP, Rubin B, Ordog T, Rijn M van de, Call J, Mühlenberg T, Fletcher JA, Bauer S. Genomic aberrations in cell cycle genes predict progression of KIT-mutant gastrointestinal stromal tumors (GISTs). Clinical Sarcoma Research BioMed Central; 2019;9(1):1–15. doi: [10.1186/s13569-019-0112-7](#)
8. Joensuu H, Roberts PJ, Sarlomo-Rikala M, Andersson LC, Tervahartiala P, Tuveson D, Silberman SL, Capdeville R, Dimitrijevic S, Druker B, Demetri GD. Effect of the Tyrosine Kinase Inhibitor STI571 in a Patient with a Metastatic Gastrointestinal Stromal Tumor. New England Journal of Medicine 2001 Apr;344(14):1052–1056. doi: [10.1056/NEJM200104053441404](#)
9. Demetri GD, Mehren M von, Blanke CD, Van den Abbeele AD, Eisenberg B, Roberts PJ, Heinrich MC, Tuveson DA, Singer S, Janicek M, Fletcher JA, Silverman SG, Silberman SL, Capdeville R, Kiese B, Peng B, Dimitrijevic S, Druker BJ, Corless C, Fletcher CDM, Joensuu H. Efficacy and Safety of Imatinib Mesylate in Advanced Gastrointestinal Stromal Tumors. New England Journal of Medicine 2002 Aug;347(7):472–480. doi: [10.1056/NEJMoa020461](#)
10. Verweij J, Casali PG, Zalcberg J, LeCesne A, Reichardt P, Blay JY, Issels R, Van Oosterom A, Hogendoorn PCW, Van Glabbeke M, Bertulli R, Judson I. Progression-free survival in gastrointestinal stromal tumours with high-dose imatinib: Randomised trial. Lancet 2004;364(9440):1127–1134. PMID:[15451219](#)
11. Cohen P. Protein kinases — the major drug targets of the twenty-first century? Nature Reviews Drug Discovery 2002 Apr;1(4):309–315. doi: [10.1038/nrd773](#)
12. Heinrich MC. Imatinib treatment of metastatic GIST: Don't stop (believing). The Lancet

Oncology Elsevier Ltd; 2010;11(10):910–911. PMID:[20864404](#)

13. Cohen P, Cross D, Jänne PA. Kinase drug discovery 20 years after imatinib: progress and future directions. *Nature Reviews Drug Discovery* Springer US; 2021;20(7):551–569. PMID:[34002056](#)

14. Casali PG, Blay JY, Abecassis N, Bajpai J, Bauer S, Biagini R, Bielack S, Bonvalot S, Boukovinas I, Bovee JVMG, Boye K, Brodowicz T, Buonadonna A, De Álava E, Dei Tos AP, Del Muro XG, Dufresne A, Eriksson M, Fedenko A, Ferraresi V, Ferrari A, Frezza AM, Gasperoni S, Gelderblom H, Gouin F, Grignani G, Haas R, Hassan AB, Hindi N, Hohenberger P, Joensuu H, Jones RL, Jungels C, Jutte P, Kasper B, Kawai A, Kopeckova K, Krákorová DA, Le Cesne A, Le Grange F, Legius E, Leithner A, Lopez-Pousa A, Martin-Broto J, Merimsky O, Messiou C, Miah AB, Mir O, Montemurro M, Morosi C, Palmerini E, Pantaleo MA, Piana R, Piperno-Neumann S, Reichardt P, Rutkowski P, Safwat AA, Sangalli C, Sbaraglia M, Scheipl S, Schöffski P, Sleijfer S, Strauss D, Strauss SJ, Hall KS, Trama A, Unk M, Sande MAJ van de, Graaf WTA van der, Houdt WJ van, Frebourg T, Gronchi A, Stacchiotti S. Gastrointestinal stromal tumours: ESMO–EURACAN–GENTURIS Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology European Society for Medical Oncology*; 2022;33(1):20–33. PMID:[34560242](#)

15. Eisenberg BL, Harris J, Blanke CD, Demetri GD, Heinrich MC, Watson JC, Hoffman JP, Okuno S, Kane JM, Mehren M von. Phase II trial of neoadjuvant/adjuvant imatinib mesylate (IM) for advanced primary and metastatic/recurrent operable gastrointestinal stromal tumor (GIST): Early results of RTOG 0132/ACRIN 6665. *Journal of Surgical Oncology* 2009 Jan;99(1):42–47. PMID:[18942073](#)

16. Wang D, Zhang Q, Blanke CD, Demetri GD, Heinrich MC, Watson JC, Hoffman JP, Okuno S, Kane JM, Von Mehren M, Eisenberg BL. Phase II trial of neoadjuvant/adjuvant imatinib mesylate for advanced primary and metastatic/recurrent operable gastrointestinal stromal tumors: Long-term follow-up results of radiation therapy oncology group 0132. *Annals of Surgical Oncology* 2012;19(4):1074–1080. PMID:[22203182](#)

17. Rutkowski P, Gronchi A, Hohenberger P, Bonvalot S, Schöffski P, Bauer S, Fumagalli E, Nyckowski P, Nguyen B-P, Kerst JM, Fiore M, Bylina E, Hoiczky M, Cats A, Casali PG, Le Cesne A, Treckmann J, Stoeckle E, Wilt JohannesHW de, Sleijfer S, Tielen R, Graaf W van der, Verhoef C, Coevorden F van. Neoadjuvant Imatinib in Locally Advanced Gastrointestinal Stromal Tumors (GIST): The EORTC STBSG Experience. *Annals of Surgical Oncology* 2013 Sep;20(9):2937–2943. PMID:[23760587](#)

18. Casali PG, Le Cesne A, Velasco AP, Kotasek D, Rutkowski P, Hohenberger P, Fumagalli E, Judson IR, Italiano A, Gelderblom H, Penel N, Hartmann JT, Duffaud F, Goldstein D, Martin-Broto J, Gronchi A, Wardelmann E, Marréaud S, Zalcberg JR, Litière S, Blay J-Y. Final analysis of the randomized trial on imatinib as an adjuvant in localized gastrointestinal stromal tumors (GIST) from the EORTC Soft Tissue and Bone Sarcoma Group (STBSG), the Australasian Gastro-Intestinal Trials Group (AGITG), UNICANCER, French Sar. *Annals of Oncology* 2021 Apr;32(4):533–541. doi: [10.1016/j.annonc.2021.01.004](#)

19. Fletcher CDM, Berman JJ, Corless C, Gorstein F, Lasota J, Longley BJ, Miettinen M, O’Leary TJ, Remotti H, Rubin BP, Shmookler B, Sobin LH, Weiss SW. Diagnosis of gastrointestinal stromal tumors: A consensus approach. *Human Pathology* 2002;33(5):459–465. PMID:[12094370](#)

20. Miettinen M, Lasota J. Gastrointestinal stromal tumors: Pathology and prognosis at different sites. *Seminars in Diagnostic Pathology* 2006;23(2):70–83. PMID:[17193820](#)

21. Joensuu H. Risk stratification of patients diagnosed with gastrointestinal stromal tumor. *Human Pathology* 2008 Oct;39(10):1411–1419. PMID:[18774375](#)

22. Hornick JL, Fletcher CDM. Immunohistochemical Staining for KIT (CD117) in Soft Tissue Sarcomas Is Very Limited in Distribution. *American Journal of Clinical Pathology* 2002 Feb;117(2):188–193. doi: [10.1309/LX9U-F7P0-UWDH-8Y6R](#)

23. West RB, Corless CL, Chen X, Rubin BP, Subramanian S, Montgomery K, Zhu S, Ball CA,



Nielsen TO, Patel R, Goldblum JR, Brown PO, Heinrich MC, Rijn M van de. The Novel Marker, DOG1, Is Expressed Ubiquitously in Gastrointestinal Stromal Tumors Irrespective of KIT or PDGFRA Mutation Status. *The American Journal of Pathology* 2004 Jul;165(1):107–113. doi: [10.1016/S0002-9440\(10\)63279-8](https://doi.org/10.1016/S0002-9440(10)63279-8)

24. Espinosa I, Lee C-H, Kim MK, Rouse B-T, Subramanian S, Montgomery K, Varma S, Corless CL, Heinrich MC, Smith KS, Wang Z, Rubin B, Nielsen TO, Seitz RS, Ross DT, West RB, Cleary ML, Rijn M van de. A Novel Monoclonal Antibody Against DOG1 is a Sensitive and Specific Marker for Gastrointestinal Stromal Tumors. *American Journal of Surgical Pathology* 2008 Feb;32(2):210–218. doi: [10.1097/PAS.0b013e3181238cec](https://doi.org/10.1097/PAS.0b013e3181238cec)

25. Miettinen M, Wang ZF, Lasota J. DOG1 antibody in the differential diagnosis of gastrointestinal stromal tumors: A study of 1840 cases. *American Journal of Surgical Pathology* 2009;33(9):1401–1408. PMID:[19606013](https://pubmed.ncbi.nlm.nih.gov/19606013/)

26. Cananzi FCM, Judson I, Lorenzi B, Benson C, Mudan S. Multidisciplinary care of gastrointestinal stromal tumour: A review and a proposal for a pre-treatment classification. *European Journal of Surgical Oncology Elsevier Ltd*; 2013;39(11):1171–1178. PMID:[24063969](https://pubmed.ncbi.nlm.nih.gov/24063969/)

27. Grazzini G, Guerri S, Cozzi D, Danti G, Gasperoni S, Pradella S, Miele V. Gastrointestinal stromal tumors: relationship between preoperative CT features and pathologic risk stratification. *Tumori* 2021;107(6):556–563. PMID:[33620027](https://pubmed.ncbi.nlm.nih.gov/33620027/)

28. Polkowski M, Gerke W, Jarosz D, Nasierowska-Guttmejer A, Rutkowski P, Nowecki ZI, Ruka W, Regula J, Butruk E. Diagnostic yield and safety of endoscopicultrasound guided trucut biopsy in patients with gastric submucosal tumors: A prospective study. *Endoscopy* 2009;41(4):329–334. PMID:[19340737](https://pubmed.ncbi.nlm.nih.gov/19340737/)

29. Cananzi FCM, Ruspi L, Samà L, Renne SL, Sicoli F, Quagliuolo V. The gist of surgical margins in GIST: a narrative review. *Laparoscopic Surgery* 2021;0:0–0. doi: [10.21037/ls-20-139](https://doi.org/10.21037/ls-20-139)

30. Renne SL. How to measure your microscope's HPF. A critical guide for residents. *Pathologica* 2023;115(6):302–307. doi: [10.32074/1591-951X-900](https://doi.org/10.32074/1591-951X-900)

31. Larghi A, Fuccio L, Chiarello G, Attili F, Vanella G, Paliani GB, Napoleone M, Rindi G, Larocca LM, Costamagna G, Ricci R. Fine-needle tissue acquisition from subepithelial lesions using a forward-viewing linear echoendoscope. *Endoscopy* 2014;46(1):39–45. PMID:[24218311](https://pubmed.ncbi.nlm.nih.gov/24218311/)

32. Kobara H, Mori H, Rafiq K, Fujihara S, Nishiyama N, Chiyo T, Matsunaga T, Ayaki M, Yachida T, Kato K, Kamada H, Fujita K, Morishita A, Oryu M, Tsutsui K, Iwama H, Kushida Y, Haba R, Masaki T. Analysis of the amount of tissue sample necessary for mitotic count and Ki-67 index in gastrointestinal stromal tumor sampling. *Oncology Reports* 2015;33(1):215–222. PMID:[25405369](https://pubmed.ncbi.nlm.nih.gov/25405369/)

33. Christensen AW, Goldberg AF. Mitotic count of fine needle aspiration material of gastrointestinal stromal tumours of the stomach underestimates actual mitotic count. *Cytopathology* 2022 Jan;33(1):100–106. doi: [10.1111/cyt.13050](https://doi.org/10.1111/cyt.13050)

34. Pearl J. Causality: Models, reasoning and inference. 2nd ed. Cambridge University Press; 2009.

35. Peters J, Janzing D, Schölkopf B. Elements of causal inference: Foundations and learning algorithms. MIT Press; 2017. Available from: <https://mitpress.mit.edu/books/elements-causal-inference> ISBN:978-0-262-03731-0

36. Talts S, Betancourt M, Simpson D, Vehtari A, Gelman A. Validating Bayesian Inference Algorithms with Simulation-Based Calibration. 2018;1–19. Available from: <http://arxiv.org/abs/1804.06788>

37. Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. Visualization in Bayesian workflow. *Journal of the Royal Statistical Society Series A: Statistics in Society Blackwell Publishing Ltd*; 2019 Feb;182(2):389–402. doi: [10.1111/rssa.12378](https://doi.org/10.1111/rssa.12378)

38. Gelman A, Vehtari A, Simpson D, Margossian CC, Carpenter B, Yao Y, Kennedy L, Gabry J, Bürkner P-C, Modrák M. Bayesian Workflow. 2020 Nov; Available from:

<http://arxiv.org/abs/2011.01808>

39. McElreath R. Statistical rethinking: A Bayesian course with examples in R and Stan. CRC press; 2020. ISBN:0429639147
40. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. Stan: A probabilistic programming language. Journal of statistical software Columbia Univ., New York, NY (United States); Harvard Univ., Cambridge, MA (United States); 2017;76(1).
41. Stan Development Team. RStan: The R interface to Stan. 2022. Available from: <https://mc-stan.org/>
42. Stan Development Team. The Stan Core Library. 2019. Available from: <http://mc-stan.org/>
43. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: <https://www.R-project.org/>
44. Hoffman MD, Gelman A. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research 2014;15:1593–1623. Available from: <https://arxiv.org/abs/1111.4246>
45. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. Florida,USA: CRC press; 2013.
46. Gelman A, Vehtari A. What are the Most Important Statistical Ideas of the Past 50 Years? Journal of the American Statistical Association Taylor & Francis; 2021 Oct;116(536):2087–2097. doi: [10.1080/01621459.2021.1938081](https://doi.org/10.1080/01621459.2021.1938081)
47. Betancourt M. Diagnosing Suboptimal Cotangent Disintegrations in Hamiltonian Monte Carlo. 2016; Available from: <http://arxiv.org/abs/1604.00695>
48. Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C. Rank-Normalization, Folding, and Localization: An Improved  $\hat{R}$  for Assessing Convergence of MCMC (with Discussion). Bayesian Analysis 2021 Jun;16(2):667–718. doi: [10.1214/20-BA1221](https://doi.org/10.1214/20-BA1221)
49. Gabry J. Graphical posterior predictive checks using the bayesplot package. 2022. Available from: <https://mc-stan.org/bayesplot/articles/graphical-ppcs.html>
50. Stan Development Team. Reparameterization. 2022. Available from: <https://mc-stan.org/docs/stan-users-guide/ppcs.html>
51. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence diagnosis and output analysis for MCMC. R News 2006;6(1):7–11. Available from: <https://journal.r-project.org/archive/>
52. Akaike H. Maximum likelihood identification of gaussian autoregressive moving average models. Biometrika 1973;60(2):255–265. doi: [10.1093/biomet/60.2.255](https://doi.org/10.1093/biomet/60.2.255)
53. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research 2010;11(Dec):3571–3594.
54. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing Springer; 2017;27(5):1413–1432.
55. Watanabe S. Mathematical Theory of Bayesian Statistics. Website. 2018. Available from: [http://watanabe-www.math.dis.titech.ac.jp/users/swatanab/mt\\_bs.html](http://watanabe-www.math.dis.titech.ac.jp/users/swatanab/mt_bs.html)
56. Watanabe S. Mathematical Theory of Bayesian Statistics. CRC Press; 2018. Available from: <https://books.google.it/books?id=GqZYDwAAQBAJ> ISBN:9781482238082
57. Walsh S, MacDuff E, Wong N, Novelli M. G103 dataset for histopathological reporting of gastrointestinal stromal tumours. The Royal College of Pathologists 2020. Available from: <https://www.rcpath.org/uploads/assets/6741f729-adf6-4ee6-b06144967e2d734a/g103-Dataset-for-histopathological-reporting-of-gastrointestinal-stromal-tumours.pdf>
58. Laurini JA. Protocol for the Examination of Resection Specimens From Patients With Gastrointestinal Stromal Tumor (GIST). College of American Pathologists Cancer Protocols 2021. Available from: [https://documents.cap.org/protocols/Stomach.GIST\\_4.2.0.0.REL\\_CAPCP.pdf](https://documents.cap.org/protocols/Stomach.GIST_4.2.0.0.REL_CAPCP.pdf)
59. Laurini JA. Protocol for the Examination of Biopsy Specimens From Patients With



Gastrointestinal Stromal Tumor (GIST). College of American Pathologists Cancer Protocols 2021. Available from: [https://documents.cap.org/protocols/Stomach.GIST.Bx\\_4.2.0.0.REL\\_CAPCP.pdf](https://documents.cap.org/protocols/Stomach.GIST.Bx_4.2.0.0.REL_CAPCP.pdf)

60. Jannink I, Risberg B, Van Diest PJ, Baak JPA. Heterogeneity of mitotic activity in breast cancer. *Histopathology* 1996;29(5):421–428. PMID:8951486

61. Callegaro D, Miceli R, Bonvalot S, Ferguson P, Strauss DC, Levy A, Griffin A, Hayes AJ, Stacchiotti S, Pechoux CL, Smith MJ, Fiore M, Dei Tos AP, Smith HG, Mariani L, Wunder JS, Pollock RE, Casali PG, Gronchi A. Development and external validation of two nomograms to predict overall survival and occurrence of distant metastases in adults after surgical resection of localised soft-tissue sarcomas of the extremities: A retrospective analysis. *The Lancet Oncology Elsevier Ltd*; 2016;17(5):671–680. PMID:27068860

62. Pasquali S, Palmerini E, Quagliuolo V, Martin-Broto J, Lopez-Pousa A, Grignani G, Brunello A, Blay JY, Tendero O, Diaz-Beveridge R, Ferraresi V, Lugowska I, Infante G, Braglia L, Merlo DF, Fontana V, Marchesi E, Donati DM, Palassini E, Bianchi G, Marrari A, Morosi C, Stacchiotti S, Bagué S, Coindre JM, Dei Tos AP, Picci P, Bruzzi P, Miceli R, Casali PG, Gronchi A, Dani C, Villa C, Messina A, Rusi L, Nuzzo AM, Nuzzo C, De Paoli A, Buonadonna A, Comandone A, Boglione A, Livi L, Greto D, Riva N, Monti M, Pennacchioli E, De Pas T, Ippolito V, Ledesma P, Redondo A, Valverde C, Bratos R, Cruz J, Martinez Trufero J, Cubedo R, Sevilla I, Luna P, Lopez R, Sancho P, Bally O, Brahmi M, Ray-Coquard I, Cassier P, Marques N, Tassy L, Boudou-Rouquette P, Tlemsani C, Alexandre J, Goldwasser F, Bompas E, Rolland F, Perrin C, Talarmin M, Italiano A, Toulmonde M, Laramas M, Bay JO, Dubray-Longeras P, Rutkowski P. Neoadjuvant chemotherapy in high-risk soft tissue sarcomas: A Sarculator-based risk stratification analysis of the ISG-STs 1001 randomized trial. *Cancer* 2022;128(1):85–93. PMID:34643947

63. Cree IA, Tan PH, Travis WD, Wesseling P, Yagi Y, White VA, Lokuhetty D, Scolyer RA. Counting mitoses: SI(ze) matters! *Modern Pathology Springer US*; 2021;34(9):1651–1657. PMID:34079071

64. Thompson NC, Greenewald K, Lee K, Manso GF. The Computational Limits of Deep Learning. 2020; Available from: <http://arxiv.org/abs/2007.05558>

65. Destounis S, Seifert P, Somerville P, Murphy P, Morgan R, Arieno A, Young WL. Underestimation of papillary breast lesions by core biopsy: correlation to surgical excision. *Breast Cancer* 2014 Mar;21(2):128–134. doi: 10.1007/s12282-012-0361-2

66. Yıldızlı ÖO, Üntan İ, Demirci D. What is the consistency between the results of needle biopsy and prostatectomy specimen pathology results? A pilot study. *Turkish journal of medical sciences* 2021 Jun;51(3):1360–1364. PMID:33535735

67. Lago V, Martín B, Ballesteros E, Cárdenas-Rebollo JM, Minig L. Tumor Grade Correlation Between Preoperative Biopsy and Final Surgical Specimen in Endometrial Cancer: The Use of Different Diagnostic Methods and Analysis of Associated Factors. *International Journal of Gynecologic Cancer* 2018 Sep;28(7):1258–1263. doi: 10.1097/IGC.0000000000001304

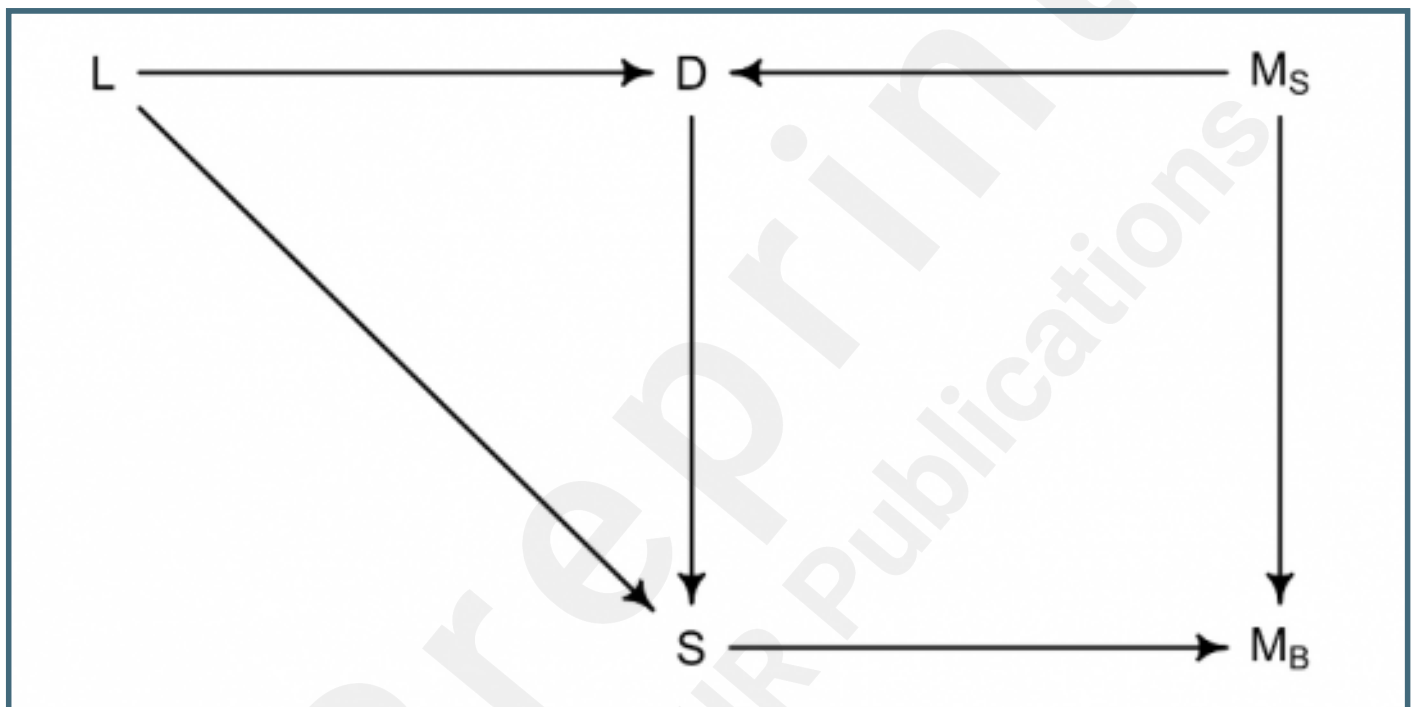
68. Dik EA, Ipenburg NA, Adriaansens SO, Kessler PA, Es RJ van, Willems SM. Poor Correlation of Histologic Parameters Between Biopsy and Resection Specimen in Early Stage Oral Squamous Cell Carcinoma. *American Journal of Clinical Pathology* 2015 Oct;144(4):659–666. doi: 10.1309/AJCPFIVHHH7Q3BLX

69. Westreich D, Greenland S. The table 2 fallacy: Presenting and interpreting confounder and modifier coefficients. *American Journal of Epidemiology* 2013;177(4):292–298. doi: 10.1093/aje/kws412

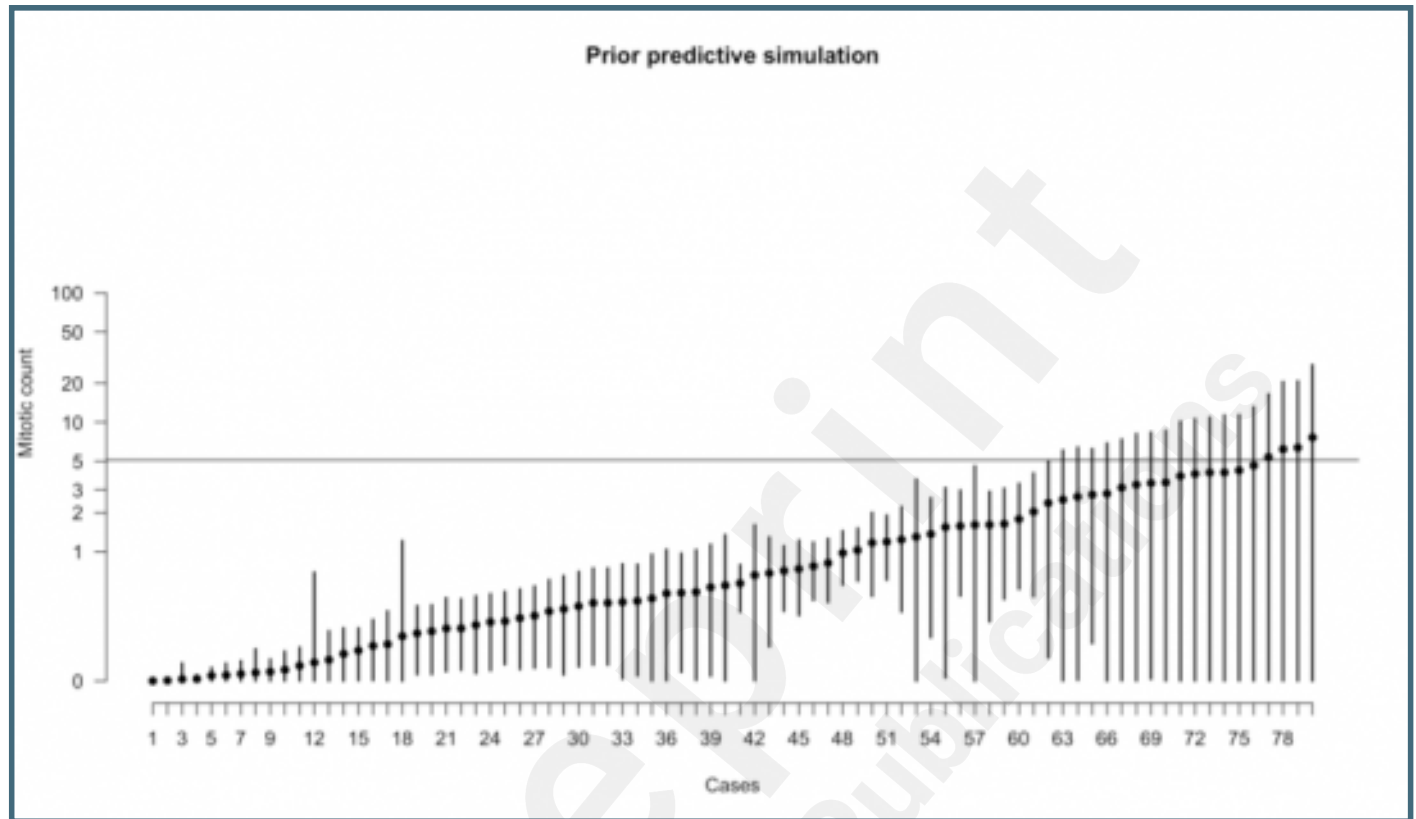
## Supplementary Files

## Figures

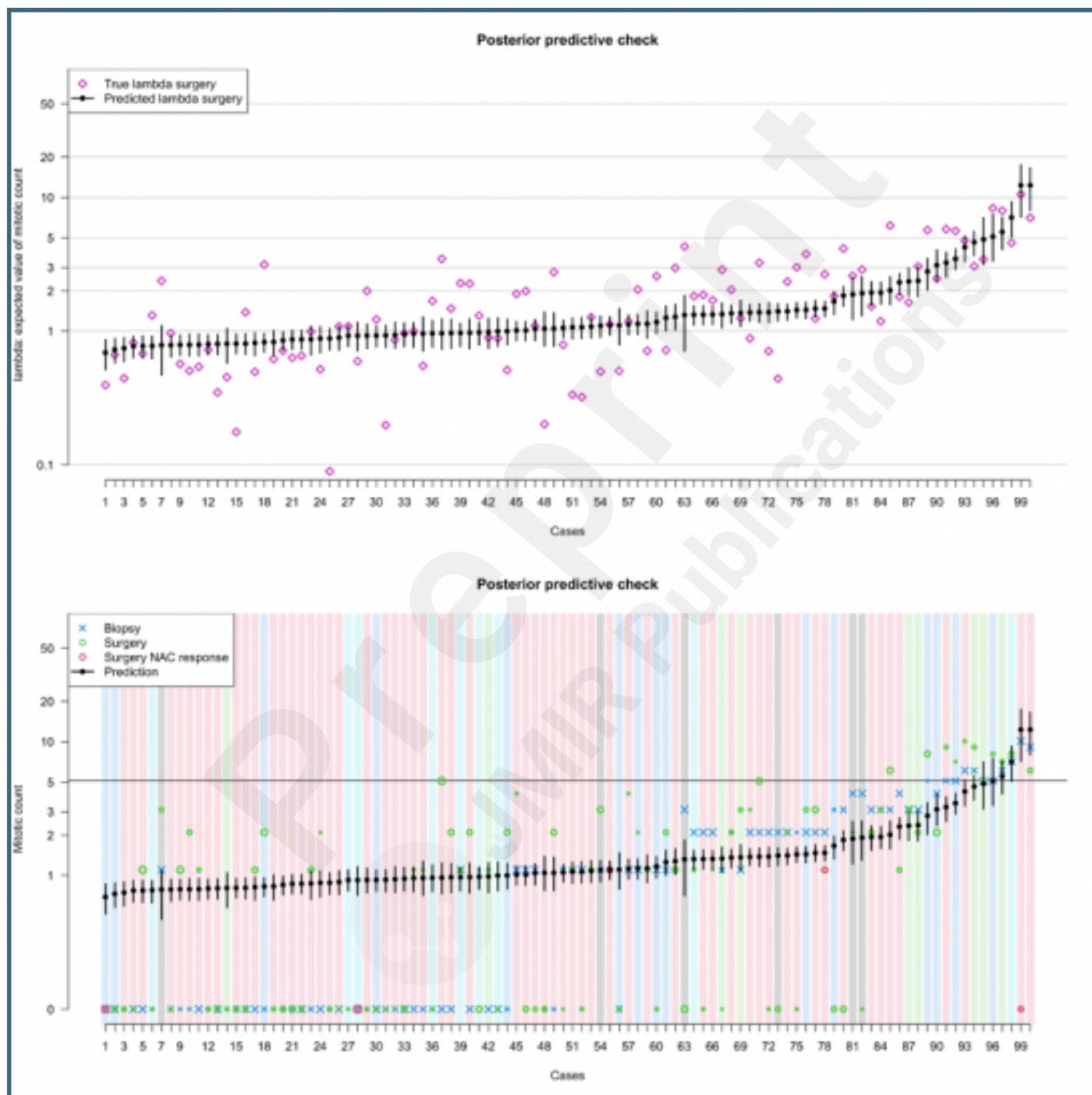
To determine the model covariates, we utilized a Directed Acyclic Graph (DAG). In this causal framework, several factors influence the mitotic count on the surgical specimen. Firstly, the dimension of the tumor impacts both the mitotic count on the biopsy and the surgical specimen. Larger tumors tend to exhibit higher mitotic activity. Secondly, the location of the tumor plays a crucial role in its growth pattern. For instance, gastric neoplasms often have more space to expand, leading to the development of symptoms with larger masses. Additionally, tumor location influences the accessibility of the biopsy site, as some sites are inherently more accessible than others. Lastly, the amount of surface area available on the biopsy directly impacts the accuracy of the mitotic count estimation. A larger surface area allows for more representative sampling of the tumor. This simplified causal model elucidates the relationship between various factors affecting GIST mitotic count. Please refer to the main text and supplementary materials for further details. In the model notation, D represents the dimension of the tumor, L denotes the location, M\_B signifies the mitotic count on the biopsy, M\_S indicates the mitotic count on the surgical specimen, and S represents the surface area of the biopsy.



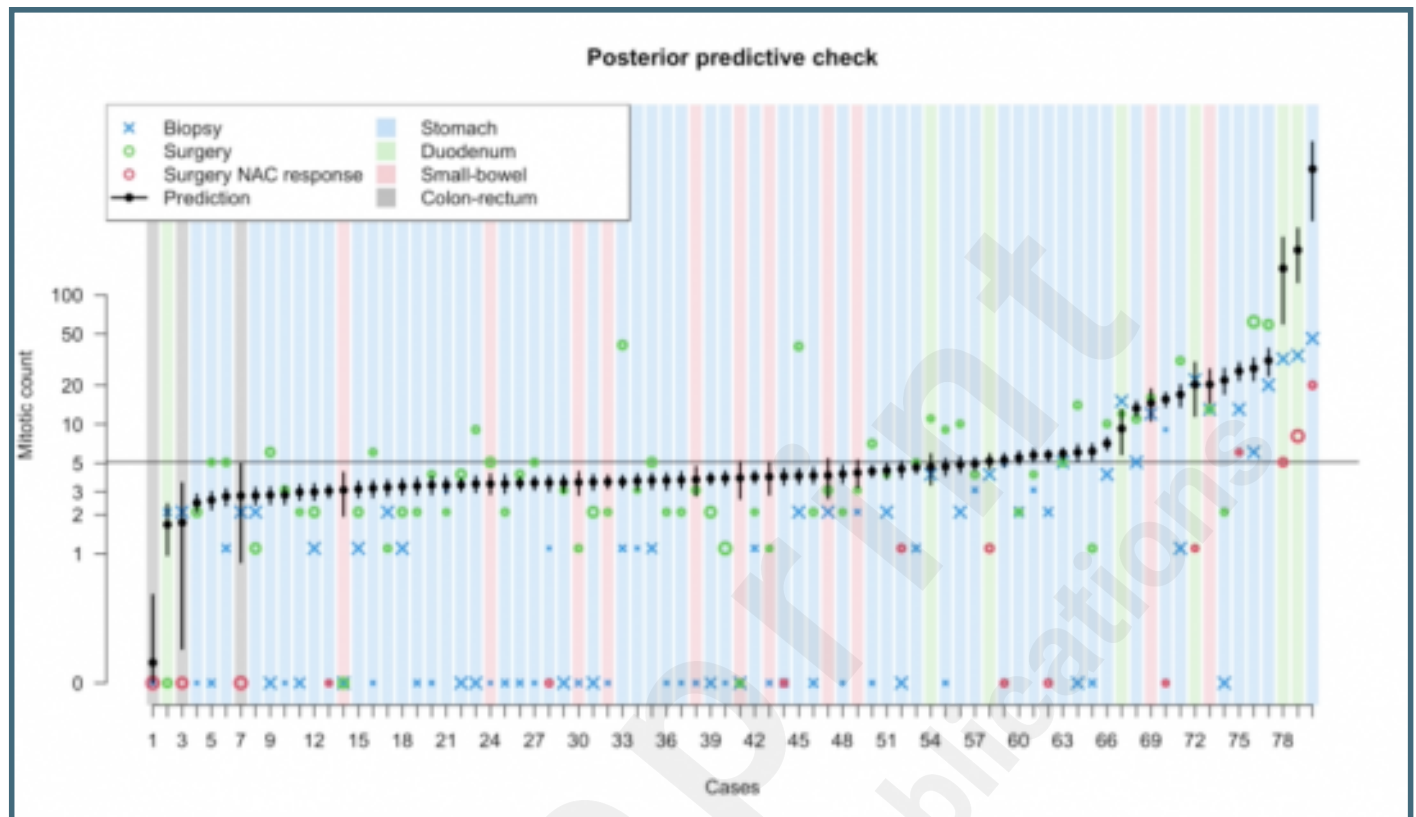
Prior predictive simulation of the  $\theta$  parameter, i.e., what the model anticipates before encountering the data. This plot displays 80 simulated cases derived from the priors. The majority of expected values indicate a very low mitotic count, aligning with real-world expectations. However, the model is not startled by higher mitotic counts, even though it anticipates encountering them in a minority of cases without prior training.



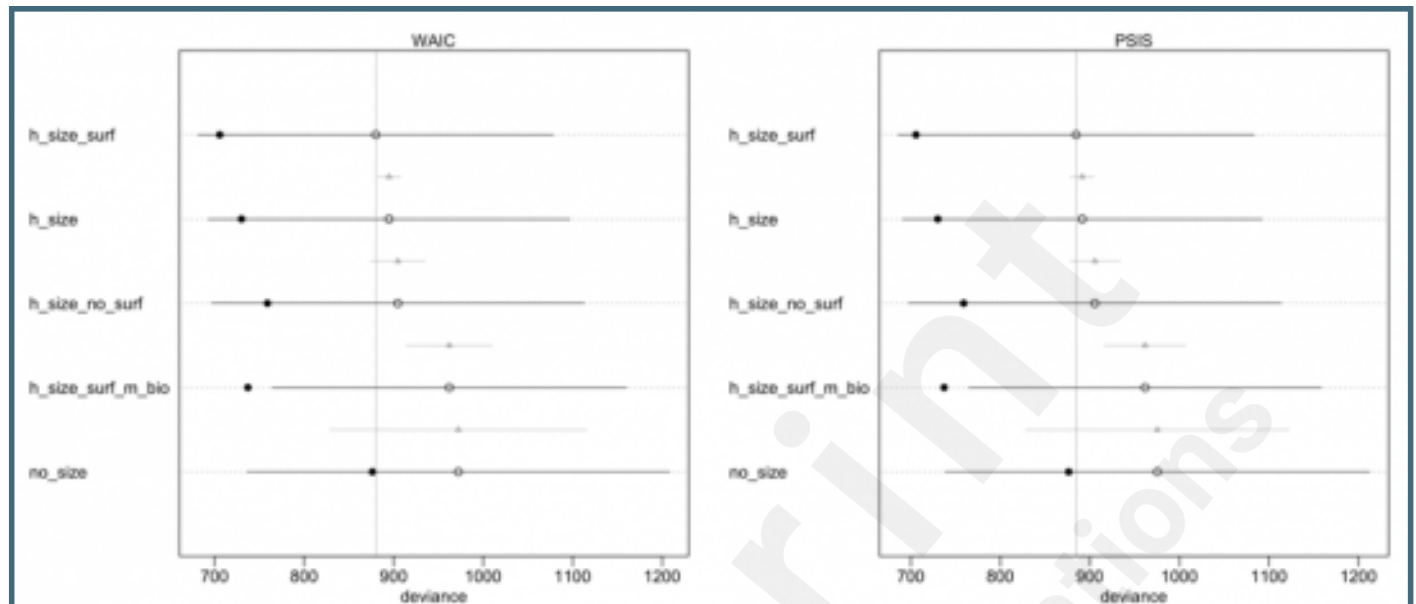
Posterior predictive simulation. The upper panel shows the imputed  $\lambda$  parameter for each sample against the ground-truth; consider the logarithmic scale on the y-axis. The lower panel shows the same imputed  $\lambda$  parameter against the fitted data. On average the mitotic count either on the surgical specimen or on the biopsy is lower than the expected value, this is due to the asymmetry of the Poisson distribution. The size of the X is proportional to the biopsy surface, the size of the empty dots is proportional to the tumor size; the color of the background corresponds to the tumor site.



Posterior predictive simulation on the true dataset. The size of the X is proportional to the biopsy surface, the size of the empty dots is proportional to the tumor size; the color of the background corresponds to the tumor site.

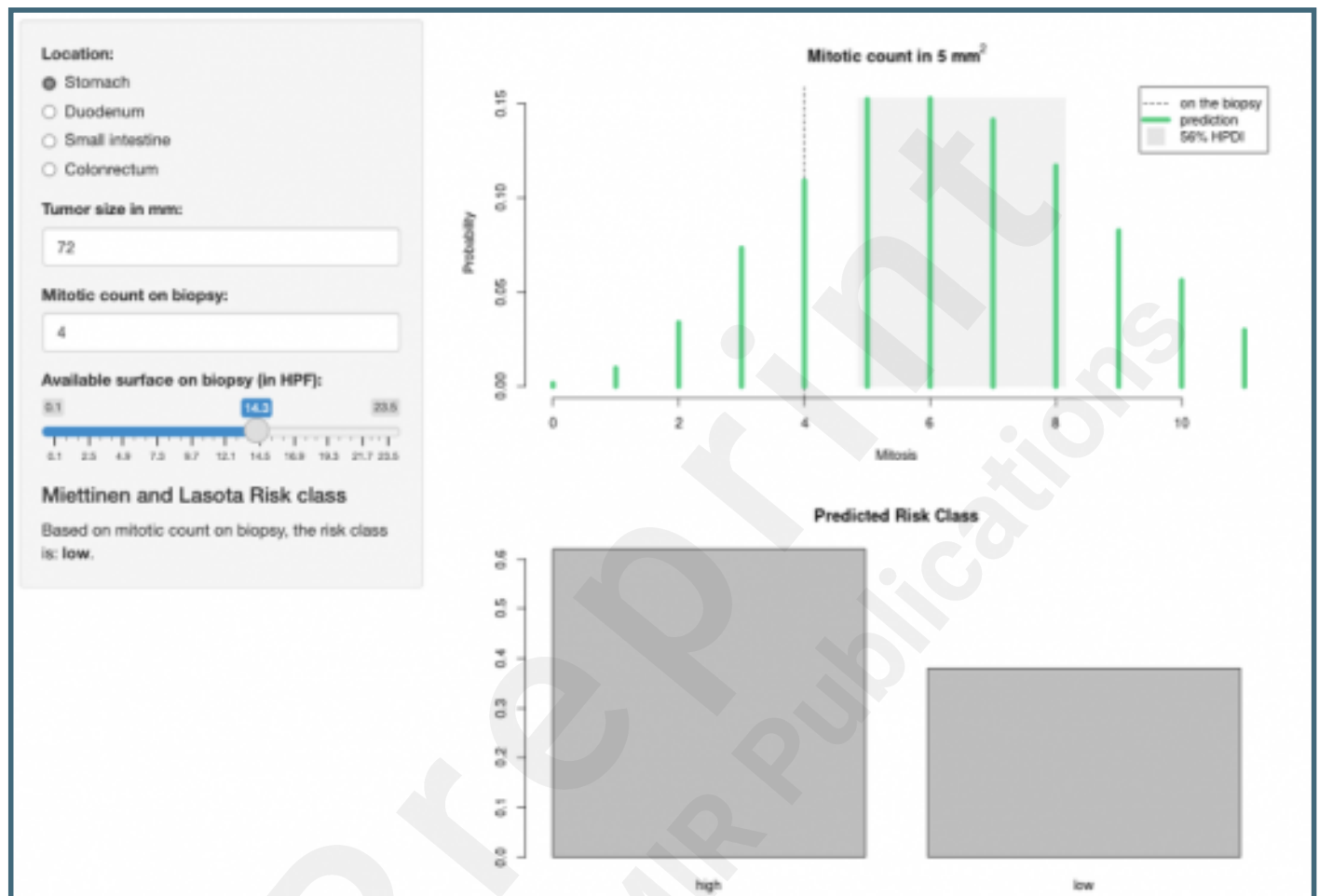


Model Selection. The vertical line indicates the mean deviance of the reference model (the one with the lowest out of sample deviance). The filled dots are values within the sample. The empty dots represent mean the out of sample deviance with the bar indicating the 89% CI; the triangle is the contrast between the model and the reference model.





The PROMETheus (PREOperative Mitosis Estimator Tool) App. The image shows a screenshot for a gastric GIST measuring 72 mm with a biopsy surface of 14.3 HPFs (about 3 mm<sup>2</sup>). With these biological characteristics the risk class is low. However, the model shows that the mitotic count is likely to be underestimated, with the most probable count on the surgical specimen predicted to be between 5 to 8 mitoses in 5 mm<sup>2</sup>. Given this predicted mitotic count distribution, it is much more probable for the risk class to be high (more than 60%). See text for further explanation.



## Multimedia Appendixes

Supplementary materials.

URL: <http://asset.jmir.pub/assets/0ccf552bbcb5ba46d3f8e070a6ed3d77.docx>

Main paper with track changes.

URL: <http://asset.jmir.pub/assets/9f91aa97a835d5b6ad1c8cd8c7afcd87.docx>

A more elaborated DAG. During optimization cycles of modeling, we greatly simplified both the DAG and the probabilistic program used. Indeed, our first modeling attempts included several latent variables (circled) as the true mitotic rate (M) and the biologic aggressiveness (U), and modeled in detail the disease presentation (including symptoms among the variables), the therapy indication and its response; although this DAG is more accurate, its complexity impairs greatly the parameter estimation efficiency. Moreover, within the domain of Bayesian network, a different approach using system of equations that literally translate the SCM in the code for the fitting might have been used, however the choice of a more simplified approach (i.e., using the M\_S as inferential target instead of as a variable) removes the need to refit the model to impute the value.

URL: <http://asset.jmir.pub/assets/94596f48e5b31d116019ea651090e0c9.png>

Prior predictive simulation for model coefficients. The coefficients for tumor size and surface (respectively ? and ?) show a wider (t) distribution compared to the biopsy count (? or ? if response to neoadjuvant therapy – NAC – is present).

URL: <http://asset.jmir.pub/assets/574c005f419b5417d02e6d1de14430cf.png>

Fit's diagnostics on the simulated dataset. The upper panel plots the number of effective samples against the R ?; the vertical gray line represents the total number of samples (half of the 8000 iterations – the first 4000 are discarded – multiplied for the 4 chains). The second panel shows the energy of the Hamiltonian, nicely following along a normal distribution (background blue line). The third panel shows a rank histogram plot (trankplot) for the log-probability of the model. Each of the four chains alternates in dominating the ranking; this indicates that the chains mixed and converged.

URL: <http://asset.jmir.pub/assets/5ac0b0786912615a9b8182227c42c24d.png>

Rank histogram plot (trankplot) for the all parameters. All the parameters have a nice mixing of the chains.

URL: <http://asset.jmir.pub/assets/278b49b162fb7ae43aaf84f81e7cb8b8.png>

Model coefficients from fitting the simulated data-set.

URL: <http://asset.jmir.pub/assets/d59daacc45e0fd9795cd197ac37ddc2b.png>

Fit's diagnostics on the real GIST cohort. The upper panel plots the number of effective samples against the R ?; the vertical gray line represents the total number of samples (half of the 4000 iterations – the first 2000 are discarded – multiplied for the 4 chains). The second panel shows the energy of the Hamiltonian, nicely following along a normal distribution (background blue line). The trankplot for the log-probability of the model shows a nice mixing of the chains.

URL: <http://asset.jmir.pub/assets/9aefc3d20e56cd58b457dbfe3e10da62.png>

Rank histogram plot (trankplot) for the all parameters. All the parameters have a nice mixing of the chains.

URL: <http://asset.jmir.pub/assets/352de2e153a51e2cf1b8c3e4d3df4f3a.png>

Model coefficients from fitting the simulated data-set.

URL: <http://asset.jmir.pub/assets/34897f99899df5a31295b18b2a2dfc6c.png>