A simulation-based sample size analysis for a glioma phase I clinical trial

Daniel Glazar[1], Solmaz Sahebjam[2], Michael Yu[3], Dung-Tsa Chen[4], Heiko Enderling[1]

[1]Integrated Mathematical Oncology, [2]Neuro-Oncology, [3]Radiation Oncology, [4]Biostatistics
[1,3,4]Moffitt Cancer Center & Research Institute, Tampa, FL, USA
[3]Center for Cancer Research, National Institutes of Health, Bethesda, MD, USA

Introduction

- Glioblastoma (GBM): aggressive cancer of the glial cells in the brain
- Low incidence (3–5 per 100k) $\sim$ low accrual
- Poor prognosis (medPFS $<$ 7 months; median OS $<$ 15 months) $\sim$ limited data per patient
- Previous work: tumor volume dynamics as an early biomarker for progression using Claret TGI model (not population approach)
- Objective: To further develop the model using the population approach and test on other clinical data. Do we have enough data?

Model specification

observation model: $y_{ij} = \overbrace{\mathsf{TS}(t_{ij};\psi_i)}^{\text{TGI model}} + \overbrace{\epsilon_{ij}}^{\text{residual error model}}$

Claret TGI model:
$$\dot{\mathsf{TS}} = \overbrace{k_{ge}\cdot \mathsf{TS}}^{\text{exponential growth}} - \overbrace{\underbrace{[t \geq 0]}_{\text{treatment indicator}}\cdot k_{\mathsf{kill}}\cdot \underbrace{e^{-\lambda t}}_{\text{evolution of resistance}}\cdot \mathsf{TS}}^{\text{treatment effect}}$$
$$\mathsf{TS}(0) = \mathsf{TS}_0$$

parameter model: $\overbrace{\ln(\psi_i)}^{\text{patient-specific parameter}} \sim \mathcal{N}(\overbrace{\ln(\mu)}^{\text{typical value}}, \overbrace{\Omega}^{\text{inter-patient heterogeneity}})$

residual error model: $\epsilon_{ij} \sim \mathcal{N}(0, a^2 + b^2\cdot \mathsf{TS}_{ij})$

survival model: $\mathsf{TS}(\mathsf{TTP}_{ij}) = \kappa_{ij}\cdot \mathsf{TS}_{\mathsf{nadir},ij}$
$\mathsf{TTP}_i \geq t_{\mathsf{nadir},ij}$

Simulating glioma patients

- Population parameters are selected to match observations from the literature.
- Tumor sizes are sampled from 2 weeks pre-treatment and every 6 weeks after treatment initiation until 714 days.
- No measurement noise is included in data generation ($a = b = 0$).
- Censoring in time occurs either at time-to-progression (TTP) or end of study (EOS = 714 days).
- Train and test sets are simulated with $N = 10, \ldots, 100$ each.
- Shown below is $N_{\mathsf{train}} = 100$ (red), $N_{\mathsf{test}} = 100$ (teal).

parameter	description	TV	IIV	unit
TS_0	tumor size at treatment initiation	16.1	1.1	cm^3
k_{ge}	growth rate	0.0123	0.84	day^{-1}
k_{kill}	kill rate	0.123	0.3	day^{-1}
λ	rate of evolution of resistance	0.02	0.3	day^{-1}
κ	relative progression threshold	1.4	0	—
LLOQ	lower limit of quantification	0.1	0	cm^3

Table 1: Population parameters used to simulate glioma patients.

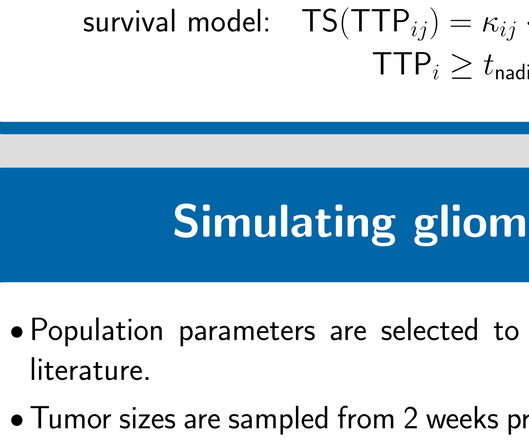


Figure 1: Patient parameter values are sampled from population distribution. Horizontal and vertical dashed lines denote minimum, median, and maximum values reported in the literature.

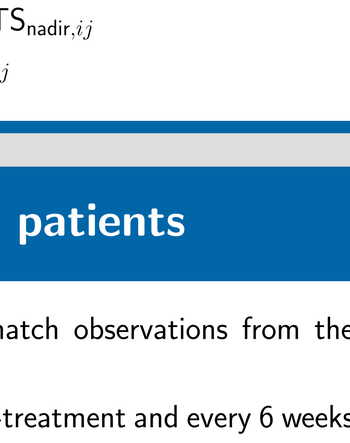
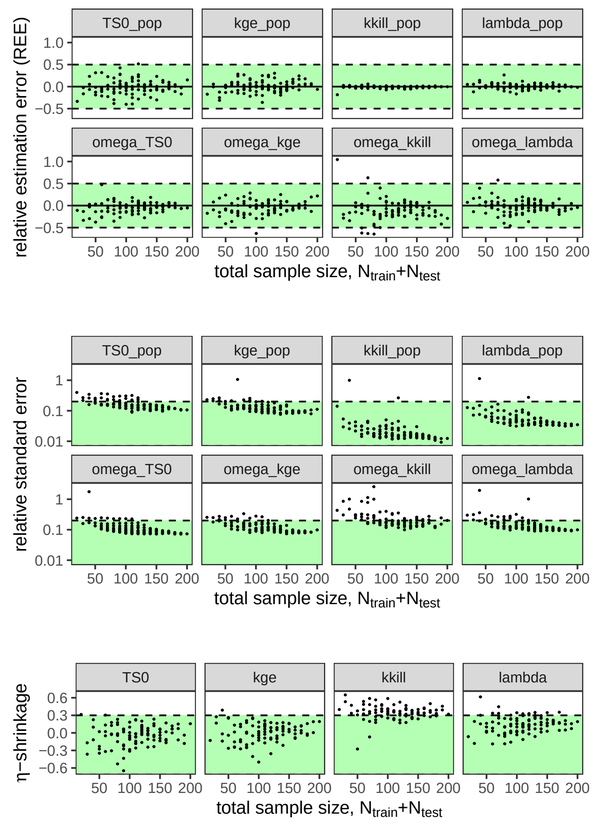


Figure 2: Spaghetti plot of tumor size (top) and Kaplan-Meier of progression-free survival (PFS; bottom).

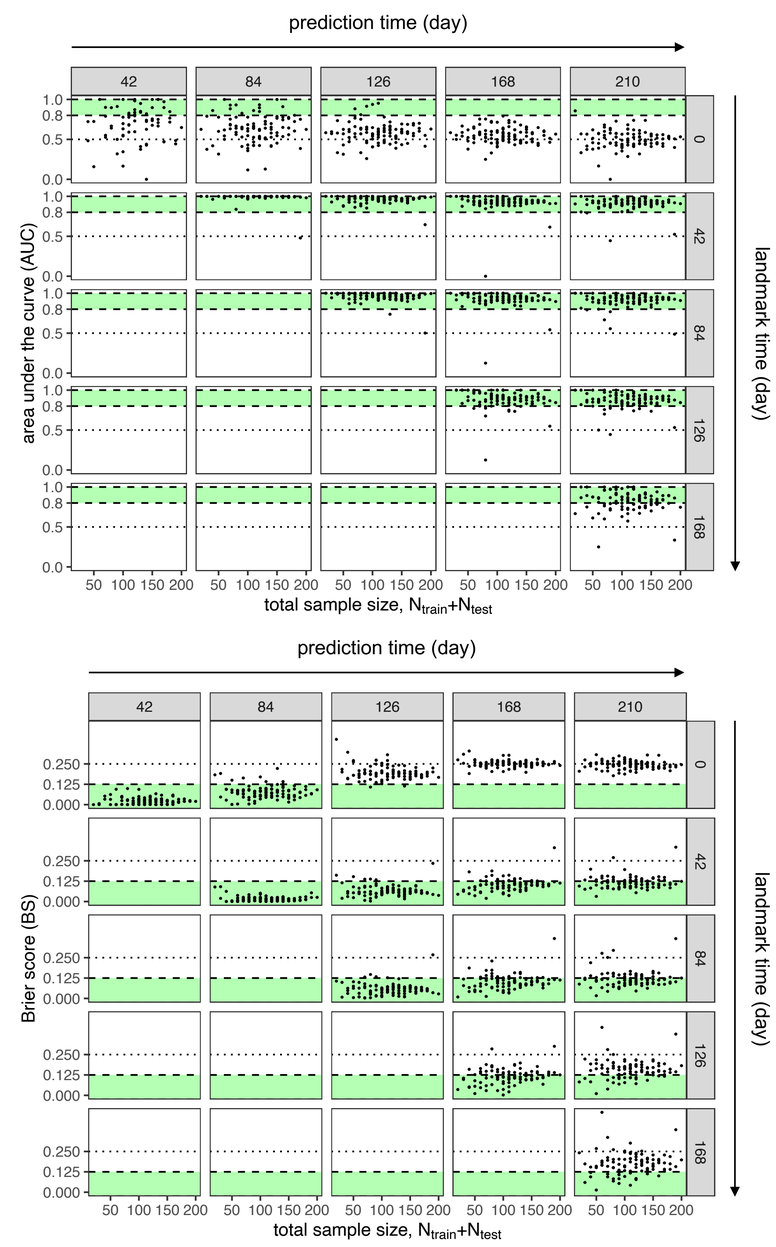
Effects of sample size on parameter estimation

- For each sample size, the developed joint model was calibrated to the train patients by estimating population parameters using SAEM, sampling the conditional distribution $p(\psi_i \mid \hat{\theta}, y_i)$ using MCMC (n_{smp}=1000), estimating EBEs, and estimating the Fisher Information Matrix (FIM). All tasks were performed using the Monolix 2023R1 software.
- Accuracy of population parameter estimates was evaluated using relative estimation error $\mathsf{REE}_k = \frac{\hat{\theta}_k - \theta_k}{\theta_k}$ (top).
- Precision of population parameter estimates was evaluated using relative standard error $\mathsf{RSE}_k = \frac{\sqrt{(\mathsf{FIM}(\hat{\theta})^{-1})_{kk}}}{\hat{\theta}_k}$ (middle).
- Data informativeness on individual parameter estimates was evaluated using η-shrinkage of the EBEs: $\eta_{\mathsf{sh},k} = 1 - \frac{\mathsf{SD}(\eta_k)}{\omega_k}$ (bottom).
- Sample size had a positive effect on RSE (Spearman's $\rho = -0.38, p = 7.22e-28$). No sample size effect was observed for neither REE nor η_{sh} ($p > 0.05$).
- Green regions denotes good values ($|\mathsf{REE}| < 0.5, \mathsf{RSE} < 0.2, \eta_{\mathsf{sh}} < 0.3$).



Effects of sample size on dynamic prediction performance

- After population parameter estimation using the train set, TTP was predicted for each test patient by sampling the conditional distribution using MCMC ($n_{\mathsf{smp}} = 1000$) across landmark times and time horizons.
- Area under the ROC curve (AUC, left) and Brier score (BS, right) were used to evaluate predictive discrimination and calibration, respectively.
- Good discrimination (AUC ≥ 0.8, green regions on left) and calibration (BS ≤ 0.125, green regions on right) were achieved for landmark times $42 \leq t \leq 84$ across all sample sizes.
- Sample size and AUC were significantly correlated ($p = 0.02$). However, the correlation was very small (Spearman's $\rho = -0.06$). No effect of sample size on BS was observed (Spearman's $\rho = 0.03, p = 0.31$).



Conclusions

- Sample size had a positive effect on the precision of population parameter estimates (RSE).
- No sample size effect was observed for accuracy of population parameter estimates (REE), data informativeness (η_{sh}), nor predictive calibration (BS).
- While a statistically significant correlation between sample size and predictive discrimination (AUC) was observed (Spearman's $p < 0.05$), the effect was very small ($\rho = -0.06$).
- The minimum sample size necessary to achieve all benchmark values was $N_{\mathsf{train}} = 60, N_{\mathsf{test}} = 10$.
- Future directions include: 1) running more replicates for each sample size; 2) adding noise, covariates (e.g., sex, treatment arm), and a sampling process to the model; 3) test using the wrong model (e.g., structural, parameter, residual); and finally 4) applying the developed model to real clinical data.