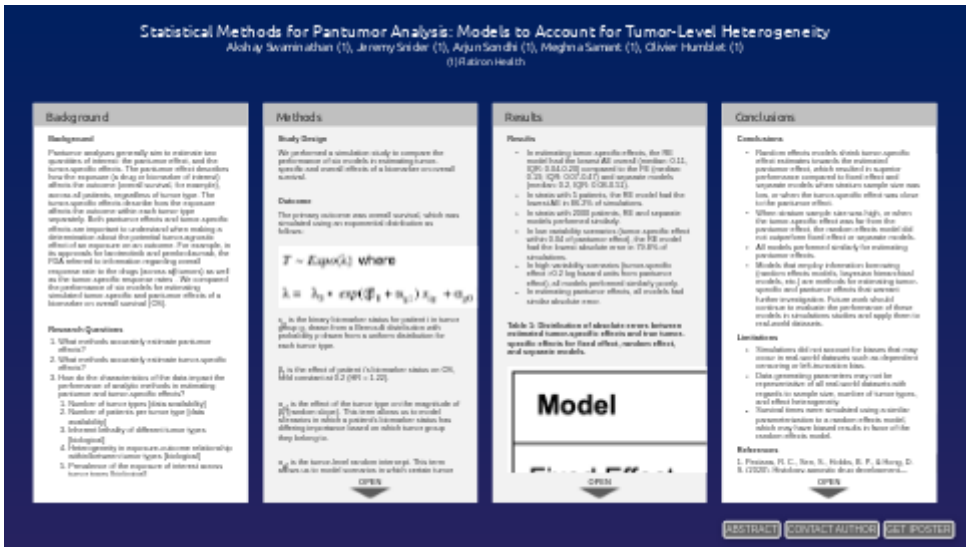
Statistical Methods for Pantumor Analysis: Models to Account for Tumor-Level Heterogeneity



Akshay Swaminathan (1), Jeremy Snider (1), Arjun Sondhi (1), Meghna Samant (1), Olivier Humblet (1)

(1) Flatiron Health

PRESENTED AT:

BACKGROUND

Background

Pantumor analyses generally aim to estimate two quantities of interest: the pantumor effect, and the tumor-specific effects. The pantumor effect describes how the exposure (a drug or biomarker of interest) affects the outcome (overall survival, for example), across all patients, regardless of tumor type. The tumor-specific effects describe how the exposure affects the outcome within each tumor type separately. Both pantumor effects and tumor-specific effects are important to understand when making a determination about the potential tumor-agnostic effect of an exposure on an outcome. For example, in its approvals for larotrectinib and pembrolizumab, the FDA referred to information regarding overall response rate to the drugs (across all tumors) as well as the tumor-specific response rates[1]. We compared the performance of six models for estimating simulated tumor-specific and pantumor effects of a biomarker on overall survival (OS).

Research Questions

1. What methods accurately estimate pantumor effects?
2. What methods accurately estimate tumor-specific effects?
3. How do the characteristics of the data impact the performance of analytic methods in estimating pantumor and tumor-specific effects?
    1. Number of tumor types [data availability]
    2. Number of patients per tumor type [data availability]
    3. Inherent lethality of different tumor types [biological]
    4. Heterogeneity in exposure-outcome relationship within/between tumor types [biological]
    5. Prevalence of the exposure of interest across tumor types [biological]

METHODS

Study Design

We performed a simulation study to compare the performance of six models in estimating tumor-specific and overall effects of a biomarker on overall survival.

Outcome

The primary outcome was overall survival, which was simulated using an exponential distribution as follows:

$$T \sim Expo(\lambda) \text{ where}$$

$$\lambda = \lambda_0 * exp((\beta_1 + \alpha_{g1})\, x_{ig} + \alpha_{g0})$$

x_{ig} is the binary biomarker status for patient i in tumor group g, drawn from a Bernoulli distribution with probability *p* drawn from a uniform distribution for each tumor type.

β_1 is the effect of patient i's biomarker status on OS, held constant at 0.2 (HR = 1.22).

α_{g1} is the effect of the tumor type on the magnitude of β_1 (random slope). This term allows us to model scenarios in which a patient's biomarker status has differing importance based on which tumor group they belong to.

α_{g0} is the tumor-level random intercept. This term allows us to model scenarios in which certain tumor groups have inherently better or worse prognosis than others.

Exposure

Patients' biomarker status was simulated using a bernoulli distribution. The probability of having the biomarker was drawn from a uniform distribution separately for each tumor type.

Simulation scenarios

We varied the number of tumor types in each dataset (5, 25, or 50), the number of patients per tumor type (5, 20, 50, 200, 500, or 2000), and the variance of the distribution of tumor-specific effects for each dataset. Each permutation was run 100 times.

Statistical analysis

To each dataset, we fit:

1. separate Cox models for each tumor type
2. a fixed-effect Cox (FE) model with interaction term between tumor type and biomarker effect

3. a random-effects (RE) Cox model with random intercept for tumor type and random slope for the biomarker effect
4. a stratified Cox model
5. the sample-size (SS) alternative to the stratified Cox (Merhotra et al., 2012)[2]
6. the and minimum-risk (MR) alternative to the stratified Cox (Merhotra et al., 2012)[2]

We report the absolute error (AE) in log hazard units between the estimated and true effects. In all boxplots, the y-axis range was restricted to remove extreme outliers.

RESULTS

Results

- In estimating tumor-specific effects, the RE model had the lowest AE overall (median: 0.11, IQR: 0.04-0.20) compared to the FE (median: 0.19, IQR: 0.07-0.47) and separate models (median: 0.2, IQR: 0.08-0.51).
- In strata with 5 patients, the RE model had the lowest AE in 86.2% of simulations.
- In strata with 2000 patients, RE and separate models performed similarly.
- In low variability scenarios (tumor-specific effect within 0.04 of pantumor effect), the RE model had the lowest absolute error in 79.8% of simulations.
- In high variability scenarios (tumor-specific effect >0.2 log hazard units from pantumor effect), all models performed similarly poorly.
- In estimating pantumor effects, all models had similar absolute error.

Table 1: Distribution of absolute errors between estimated tumor-specific effects and true tumor-specific effects for fixed effect, random effect, and separate models.

Model	Min	1st Quartile	Median	3rd Quartile	Max
Fixed Effect	0.00	0.07	0.19	0.47	62.92
Random Effect	0.00	0.04	0.11	0.25	2.93
Separate	0.00	0.08	0.20	0.51	24.38

Figure 1: Distribution of tumor-specific effects for fixed effect, random effect, and separate models, as well as the true distribution of tumor-specific effects across all simulations. The percentage of simulations where a given model's absolute error was lowest is shown.

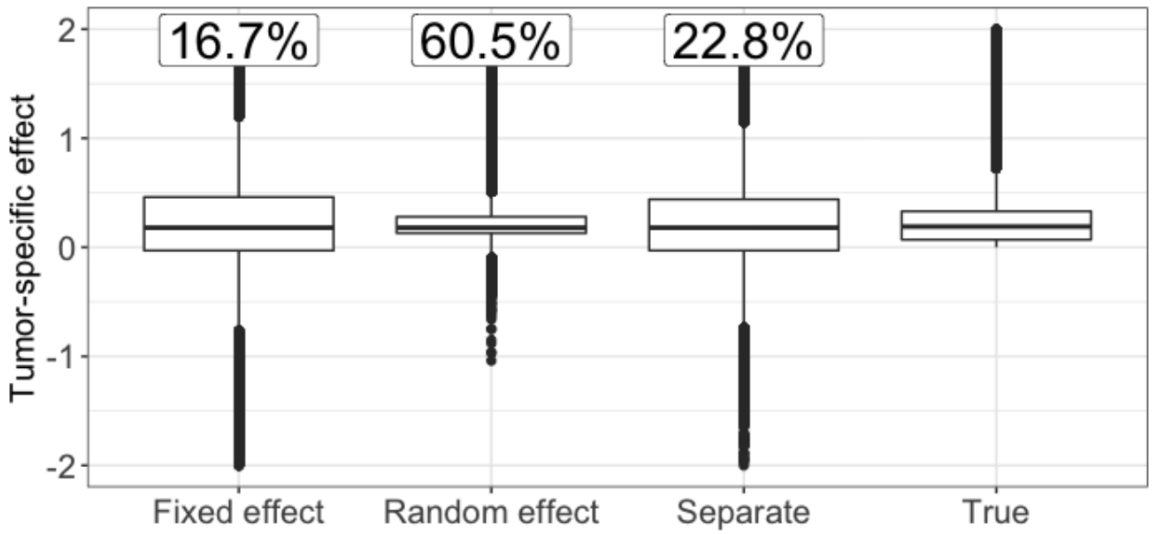


Figure 2: Absolute errors for estimated tumor-specific effects for fixed effect, random effect, and separate models, stratified by tumor type sample size. The percentage of simulations where a given model's absolute error was lowest is shown.

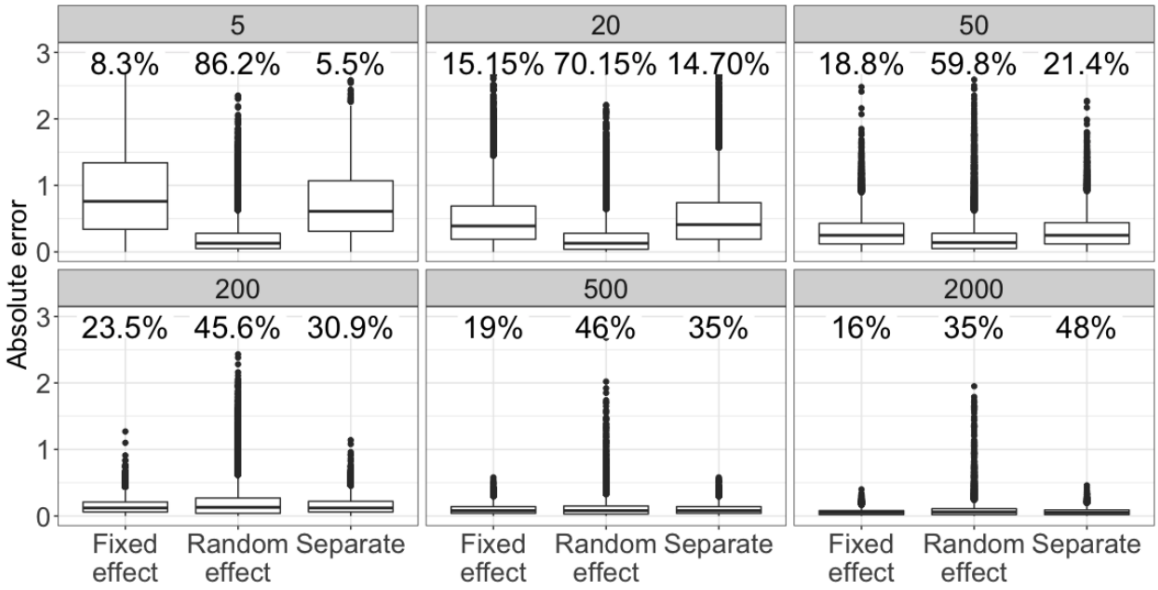


Figure 3: Absolute errors for estimated tumor-specific effects for fixed effect, random effect, and separate models, stratified by the magnitude of tumor-specific effects in each dataset. In units of absolute log hazard from the pantumor effect, "lowest" corresponds to <0.04, "low" corresponds to 0.04-0.12, "high" corresponds to 0.12-0.2, and "highest" corresponds to >0.2. The percentage of simulations where a given model's absolute error was lowest is shown.

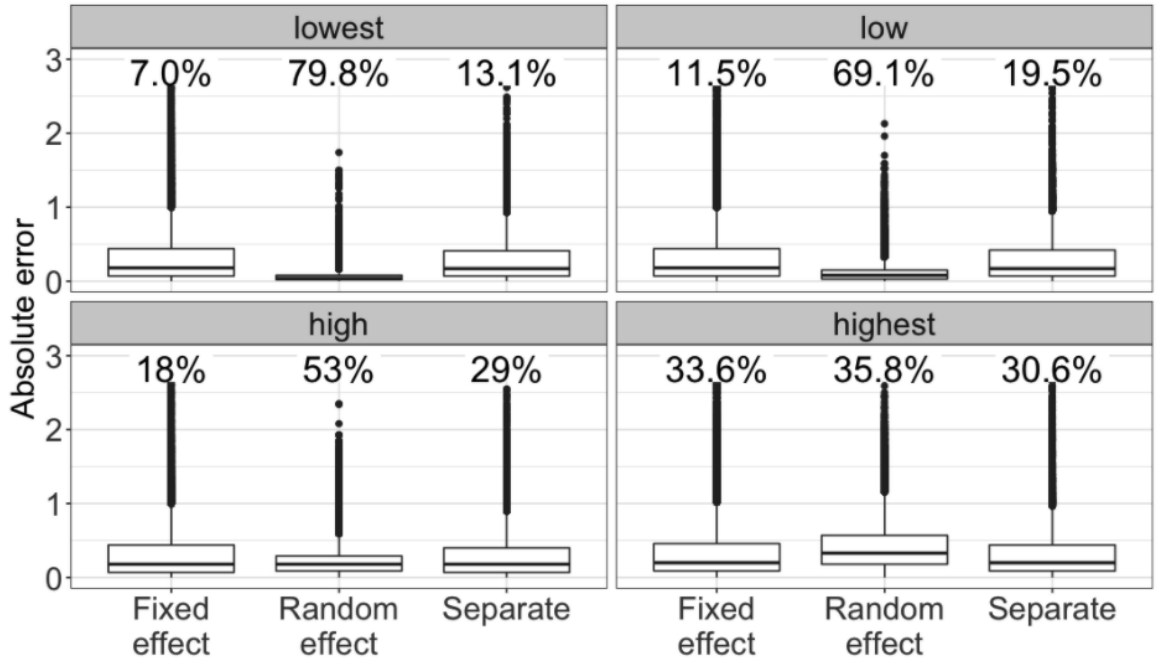


Figure 4: For tumor type strata where sample size was 20 (A) or 500 (B), absolute errors for estimated tumor-specific effects for fixed effect, random effect, and separate models, stratified by the magnitude of tumor-specific effects in each dataset are shown. In units of absolute log hazard from the pantumor effect, "lowest" corresponds to <0.04, "low" corresponds to 0.04-0.12, "high" corresponds to 0.12-0.2, and "highest" corresponds to >0.2. The percentage of simulations where a given model's absolute error was lowest is shown.

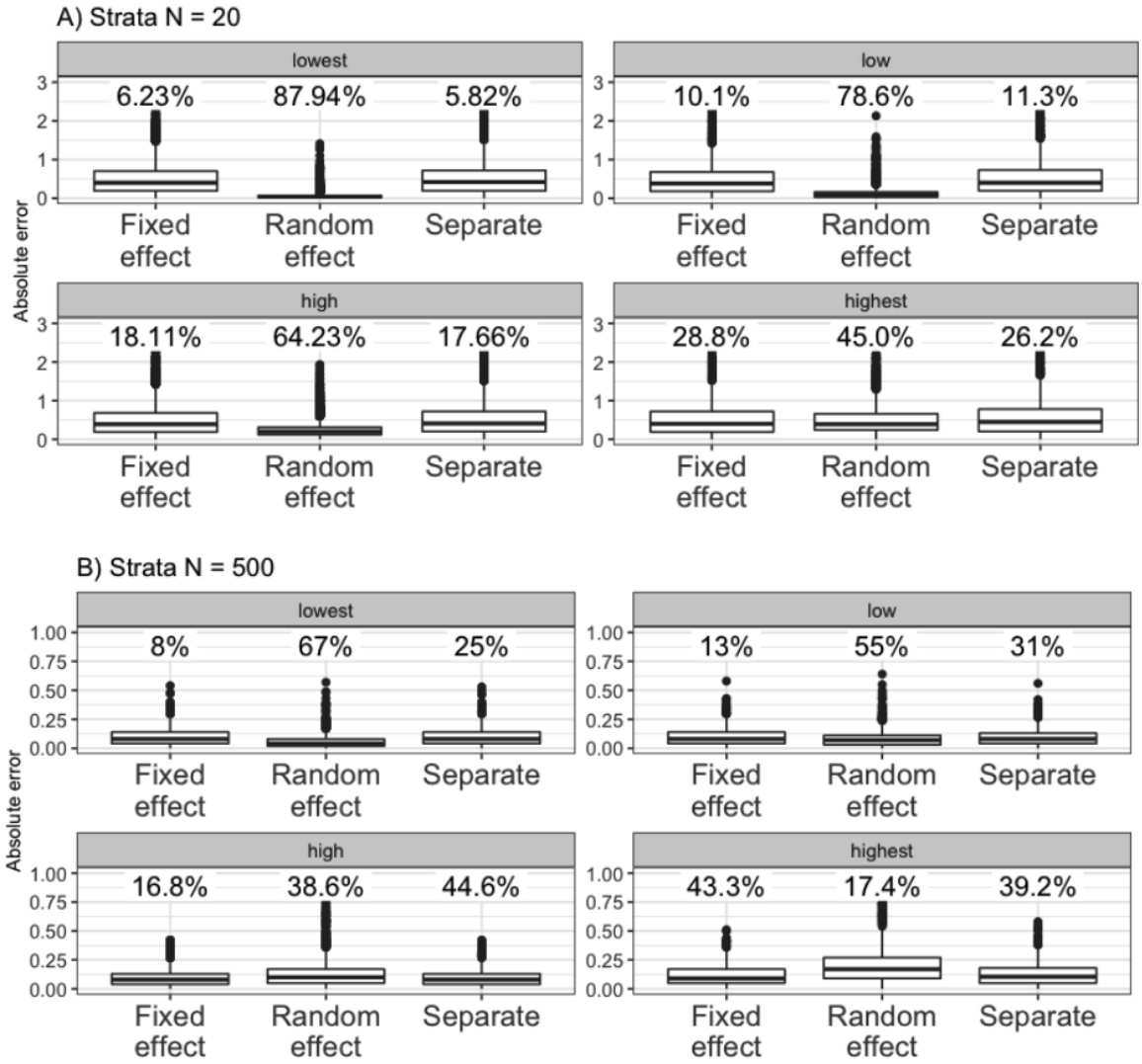
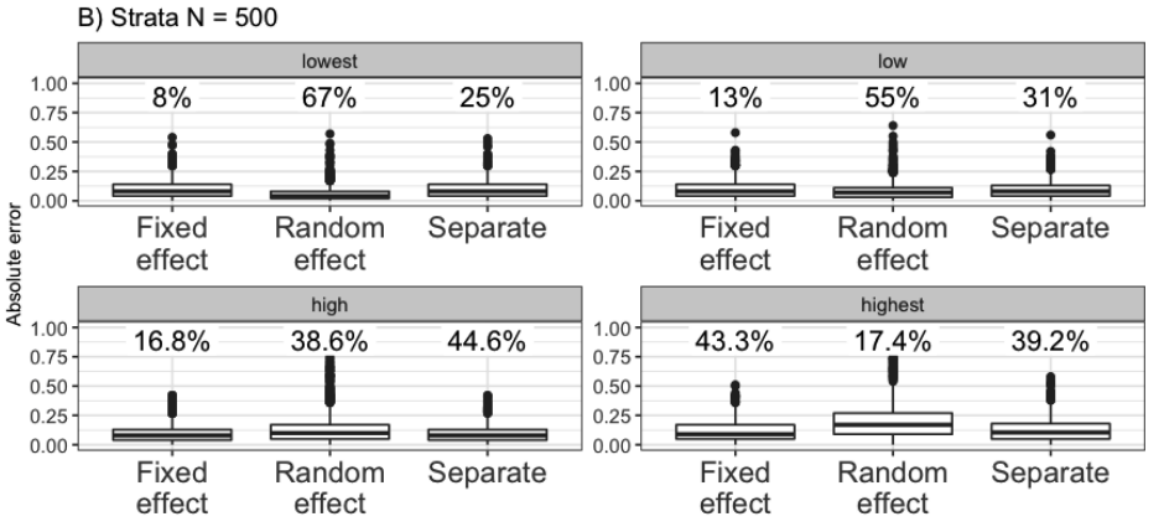


Figure 5: Distribution of pantumor effects for RE, stratified, and Merhotra models. The true pantumor effect was held constant at 0.2. The percentage of simulations where a given model's error was lowest is shown, as well as the median absolute error.



CONCLUSIONS

Conclusions

- Random effects models shrink tumor-specific effect estimates towards the estimated pantumor effect, which resulted in superior performance compared to fixed effect and separate models when stratum sample size was low, or when the tumor-specific effect was close to the pantumor effect.
- When stratum sample size was high, or when the tumor-specific effect was far from the pantumor effect, the random effects model did not outperform fixed effect or separate models.
- All models performed similarly for estimating pantumor effects.
- Models that employ information borrowing (random effects models, bayesian hierarchical models, etc.) are methods for estimating tumor-specific and pantumor effects that warrant further investigation. Future work should continue to evaluate the performance of these models in simulations studies and apply them to real-world datasets.

Limitations

- Simulations did not account for biases that may occur in real-world datasets such as dependent censoring or left-truncation bias.
- Data generating parameters may not be representative of all real-world datasets with regards to sample size, number of tumor types, and effect heterogeneity.
- Survival times were simulated using a similar parameterization to a random effects model, which may have biased results in favor of the random effects model.

References

1. Pestana, R. C., Sen, S., Hobbs, B. P., & Hong, D. S. (2020). Histology-agnostic drug development—considering issues beyond the tissue. Nature Reviews Clinical Oncology, 17(9), 555-568.

2. Mehrotra, D. V., Su, S. C., & Li, X. (2012). An efficient alternative to the stratified cox model analysis. Statistics in medicine, 31(17), 1849-1856.

ABSTRACT

Objectives

Recent tumor-agnostic regulatory approvals indicate growing interest in estimating effects of drugs and biomarkers on outcomes in specific tumor types (TTs) and across all TTs (pantumor effects), which can be challenging due to tumor-level differences in prognosis and data availability. We compared the performance of six models for estimating simulated tumor-specific and pantumor effects of a biomarker on overall survival (OS).

Methods

Patient survival times were simulated based on patients' biomarker status, tumor type, and the pantumor and tumor-specific effects of the biomarker on OS. We varied the number of TTs in each dataset (5-50), and the number of patients per TT (5-2000). Each permutation was run 100 times.

To each dataset, we fit separate Cox models for each TT, a fixed-effect Cox (FE) model, a random-effects (RE) Cox model, a stratified Cox model, and the sample-size (SS) and minimum-risk (MR) alternatives to the stratified Cox (Merhotra et al., 2012). We report the absolute error (AE) comparing the estimated and true effects (log hazard).

Results

In estimating tumor-specific effects, the RE model had the lowest AE overall (median: 0.11, IQR: 0.04-0.20) compared to the FE (median: 0.19, IQR: 0.07-0.47) and separate models (median: 0.2, IQR: 0.08-0.51). In TTs with 5 patients, the RE model had the lowest AE in 86.2% of simulations. In TTs with 2000 patients, RE and separate models performed similarly. In low variability scenarios (tumor-specific effect within 0.04 of pantumor effect), the RE model had the lowest AE in 79.8% of simulations. In high variability scenarios (tumor-specific effect >0.2 from pantumor effect), all models performed similarly. In estimating pantumor effects, all models had similar AE.

Conclusions

RE models performed favorably for estimating tumor-specific effects, and all models performed similarly for estimating pantumor effects.