Hack-a-Stat 2026

Kaplan-Meier Analysis and Simulation-Based Power Assessment for Multi-Indication Cancer Drug Development

1. Overview and Objectives

This report consists of two connected but conceptually distinct analyses:

- **Exercise 1 (E1)**: An inferential analysis based on *observed clinical data* to assess whether the study drug improves 24-month survival relative to the standard of care in lung cancer.
- **Exercise 2 (E2)**: A simulation-based evaluation of *trial operating characteristics* to assess how reliably a proposed trial would demonstrate benefit under an assumed treatment effect in a different disease context.

Although the same drug appears in both exercises, the objectives, data sources, and statistical interpretations differ fundamentally. This distinction drives all modeling and inference decisions in the report.

2. Exercise 1 — Evidence from Observed Data (Lung Cancer)

2.1 Objective of Exercise 1

The goal of Exercise 1 is to determine whether the observed data provide statistical evidence that the study drug improves **24-month event-free survival** compared with the current standard of care in lung cancer.

This is a **retrospective inferential problem**, based on real patient follow-up with censoring.

2.2 Data Preparation and Structure

[Q. Carefully explore the data and identify the characteristics of the data. Thereafter, convert the data variables into a format that will be easy to analyze. Clearly explain all the steps taken while converting the data.]

The dataset contains patient-level information on:

- Recruitment date
- Event or withdrawal date
- Reason for event or censoring

For patients where both values are present it is the difference between start and end date in months. But some values were blank for event or withdrawal time .Where the event or withdrawal date is missing, we can consider the cutoff date which is 01-01-2024 because we assume that if no date is specified then the patient neither withdrew consent from the trial nor experienced an event. Means patient is event free.

From this, we constructed:

- A **time-to-event variable** (in months from recruitment)
- An **event indicator**, where events include disease progression or death
- Right-censoring for patients who withdrew or remained event-free at study cut-off

subject_id	recruitment_date	vent_or_withdrawal_dat	reason	event	last_date	time_months
1	2020-05-03	2020-10-19	Disease Pr	1	2020-10-19	5.551905388
2	2021-07-07			0	2024-01-01	29.82917214
3	2021-11-13			0	2024-01-01	25.5913272
4	2020-04-15			0	2024-01-01	44.54664915
5	2020-04-04	2022-12-30	Withdrawa	0	2022-12-30	32.85151117
6	2020-12-20			0	2024-01-01	36.36662286
7	2021-11-06	2023-07-10	Withdrawa	0	2023-07-10	20.07227332
8	2020-03-16	2021-05-29	Disease Pr	1	2021-05-29	14.4218134
9	2021-02-21			0	2024-01-01	34.29697766
10	2020-05-15	2021-09-05	Disease Pr	1	2021-09-05	15.70302234

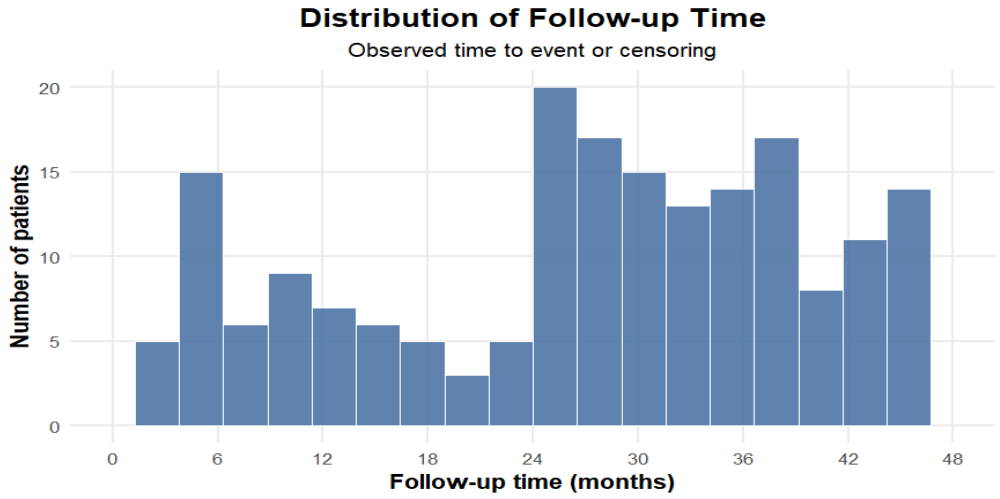
Censoring is therefore present and non-negligible, making standard proportions inappropriate.

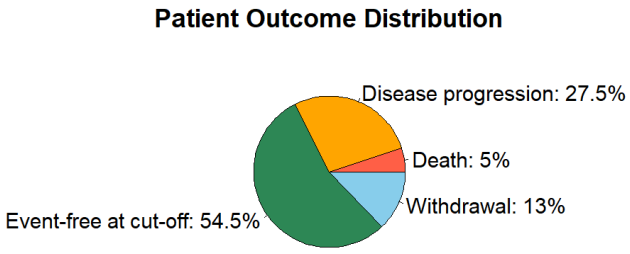
2.3 Exploratory Survival Data Assessment

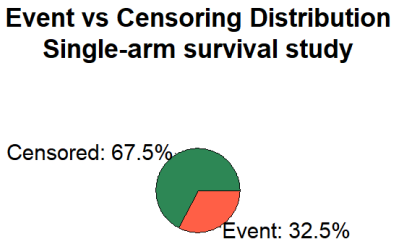
[Q.Choose an appropriate visualization plot to graphically explore the endpoint of interest. Make some exploratory remarks on the endpoint of interest.]

Before formal inference, exploratory analyses were conducted to assess:

- The balance between observed events and censoring
- The distribution of follow-up times[bi-modal]
- Whether sufficient patients remain under observation at 24 months







These checks confirmed that:

- A substantial fraction of patients experienced events
- Follow-up extended well beyond 24 months for many patients
- A large number of patients remained at risk at the 24-month landmark

This supported the use of time-to-event methods.

2.4 Choice of Estimator: Kaplan–Meier

Given the presence of right censoring and staggered entry, the **Kaplan–Meier (KM) estimator** was used to estimate survival.

Kaplan–Meier was chosen because it:

- Makes **no parametric assumptions** about the survival distribution
- Correctly accounts for censoring
- Directly estimates the estimand of interest, (S(24))

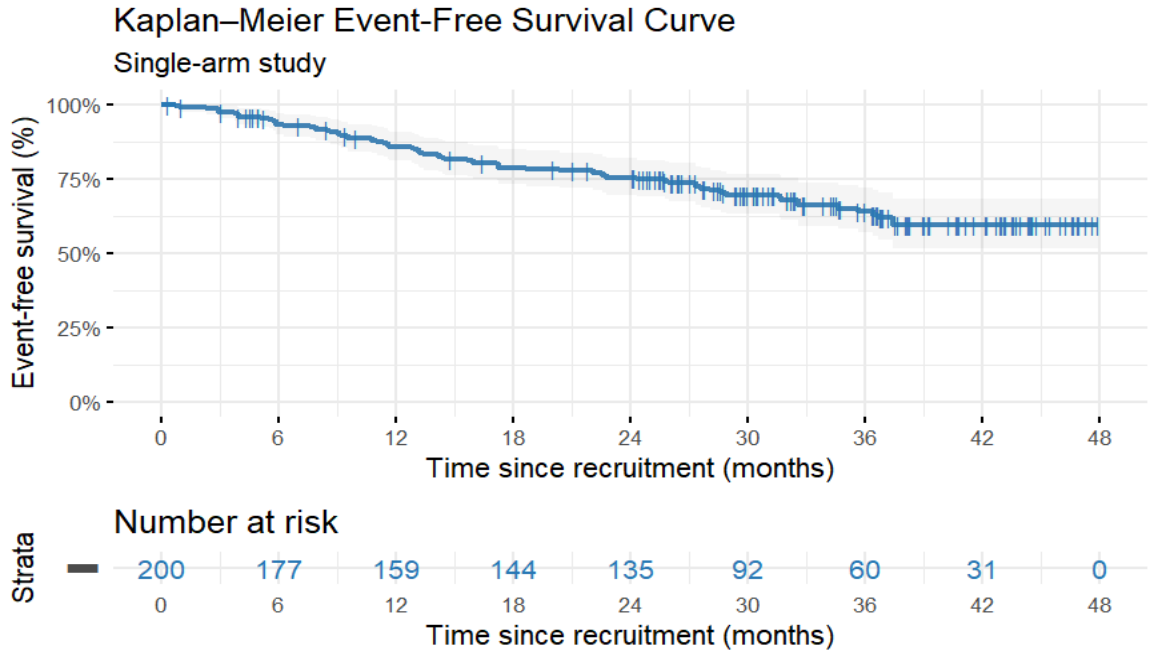
This aligns with best practice when the underlying lifetime distribution is unknown.

The observed bimodal survival lifetime likely reflects underlying patient heterogeneity and does not compromise the validity of the Kaplan–Meier–based inference, which remains robust to complex and multimodal time-to-event distribution

The approximately 30–70 split between observed events and censoring is consistent with an active treatment effect and adequate follow-up, providing sufficient event information for survival estimation while preserving a large at-risk population at later time points.

2.5 Estimation Results

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24	135	46	0.754	0.0315	0.695	0.819



At 24 months after starting treatment, approximately **75% of patients are estimated to remain event-free**, meaning they have neither experienced disease progression nor death by this time point. This estimate is based on **135 patients who were still under observation and at risk** at 24 months, providing a strong data foundation for the result.

The uncertainty around this estimate is relatively small. Even after accounting for statistical variability, the **true 24-month survival rate is very likely to lie between approximately 70% and 82%**. This range reflects normal sampling uncertainty rather than doubt about treatment effectiveness.

From a practical standpoint, this means that the observed survival benefit is not driven by a small or unstable subset of patients. Instead, it is supported by a large group of individuals followed long enough to meaningfully inform outcomes at 24 months, indicating a **robust and clinically meaningful survival profile** for the study drug.

2.6 Hypothesis Testing Framework (E1)

[Q.Conduct a rigorous statistical test to assess if the 24-month survival rate of the study drug is significantly better than the standard of care currently available. Clearly explain the statistical method used and state the assumptions required for the test to be valid.]

Let (S(24)) denote the true 24-month survival probability for patients treated with the study drug, and let ($S_0 = 0.50$) represent the benchmark survival associated with standard of care.

The hypotheses were defined as:

- **Null hypothesis**:
  (H_0: S(24) <=0.50)
- **Alternative hypothesis**:
  (H_1: S(24) > 0.50)

A **one-sample, one-sided Z-test** based on the Kaplan–Meier estimator and Greenwood's standard error was used.This test is valid as the kaplan meier estimate is asymptotically normal

This test:

- Targets fixed-time survival directly
- Accounts for censoring
- Avoids unnecessary parametric assumptions

2.7 Inference and Conclusion (Exercise 1)

The resulting test statistic was very large, with an associated p-value far below the 5% significance level.

Conclusion (Exercise 1):

The observed data provide strong statistical evidence that the study drug improves 24-month event-free survival relative to the standard of care in lung cancer.

From a practical perspective, this indicates that the observed improvement is unlikely to be due to random variation and supports the clinical relevance of the study drug in this setting.

3. Transition from Exercise 1 to Exercise 2

Exercise 1 establishes **evidence of benefit** in lung cancer based on observed data.

Exercise 2 does **not** attempt to extrapolate this result numerically. Instead, it addresses a different and forward-looking question:

If a similar magnitude of benefit were believed to hold in a different disease context, how reliable would a future trial be at demonstrating that benefit?

This distinction between **evidence generation (Exercise 1)** and **design risk assessment (Exercise 2)** is central to the analysis.

4. Exercise 2 - Simulation-Based Trial Design Evaluation

4.1 Purpose and Conceptual Framing

Exercise 2 addresses a fundamentally different objective from Exercise 1. While Exercise 1 is an inferential analysis based on observed patient data, Exercise 2 is a **trial design and operating-characteristics evaluation** conducted under explicitly stated assumptions. The treatment effect in this exercise is not estimated from data but is provided as a belief in the problem statement.

Although the same drug is considered across exercises, this is **statistically irrelevant** for Exercise 2. The disease context differs, the standard-of-care benchmark differs (40% versus 50% at 24 months), and the assumed treatment effect differs. As a result, there is no reason to expect the same survival curve, hazard structure, censoring behavior, or effect size across the two settings. Exercise 2 therefore does not attempt to extrapolate results from Exercise 1, but instead evaluates whether a proposed trial design would be capable of demonstrating benefit *if the assumed effect were true*.

4.2 Information Explicitly Provided

The problem statement specifies the following inputs, which are treated as fixed assumptions throughout the simulation study:

- **Standard-of-care 24-month survival rate**: 40%
- **Assumed study drug 24-month survival rate**: 60%
- **Annual dropout rate**: 5%
- **Patient entry times**: uniformly distributed between 0 and 24 months

No information is provided regarding the shape of the hazard function or the timing of events. Consequently, any time-to-event analysis requires explicit modeling assumptions that must be justified and kept minimal.

4.3 Choice of Lifetime Distribution

The exponential distribution was considered first as a natural baseline, as it requires the fewest quantitative assumptions and is fully specified by a single survival probability at a fixed time point. This makes it attractive for trial-design evaluation when detailed biological information is unavailable.

However, evidence from oncology studies indicates that hazards are often not constant over time. While some disease settings exhibit unimodal hazard patterns that may be modeled using log-normal or log-logistic distributions, such assumptions would require specifying the timing and shape of a hazard peak—information that is not provided in the problem statement and would therefore be speculative.

To balance realism with parsimony, the **Weibull distribution** was selected as a defensible compromise. The Weibull model represents the minimal extension of the exponential model, introducing a single shape parameter that allows for monotone time-varying hazards (either increasing or decreasing) without assuming non-monotonic behavior. Importantly, Weibull models are used here **as a sensitivity analysis framework**, not to claim biological realism, but to assess how trial operating characteristics change under plausible departures from the constant-hazard assumption while remaining consistent with the assumed 24-month survival belief.

4.4 Simulation Design

Simulated trial datasets were generated to mirror a realistic single-arm oncology study design. Each simulated trial incorporated:

- Weibull-distributed event times calibrated such that true 24-month survival equals 60%
- Exponentially distributed dropout times corresponding to a 5% annual dropout rate
- Uniform staggered entry over the first 24 months
- Administrative censoring at a fixed study end time

Three Weibull shape parameters were considered to represent distinct hazard scenarios:

Hazard Type	Shape Parameter (k)	Clinical Story	When is the patient safest?
Decreasing	k < 1 (e.g., 0.7)	**Early Failure:** "If you make it past 6 months, you're safe."	Later in the trial
Constant	k = 1	**Steady State:** "Risk is the same every day."	Equal risk always
Increasing	k > 1 (e.g., 1.3)	**Wear-Out:** "The longer you stay on, the higher the risk."	Early in the trial

Each scenario was simulated repeatedly to obtain stable estimates of trial operating characteristics.

4.5 Definition of Trial Success (Hypothesis Framework)

Each simulated trial was evaluated using the **same decision rule planned for the actual study**, ensuring alignment between trial design evaluation and eventual analysis.

The hypothesis being evaluated can be stated as follows:

- **Null hypothesis**: The study drug does **not** improve 24-month survival beyond the standard of care (i.e., the true 24-month survival rate is 40% or lower).

- **Alternative hypothesis**: The study drug **does** improve 24-month survival beyond the standard of care (i.e., the true 24-month survival rate is greater than 40%).

For each simulated trial:

- The 24-month survival probability was estimated using the Kaplan–Meier method.

- A one-sided statistical test was applied at the 5% significance level.

- A trial was declared **successful** if the analysis concluded that 24-month survival exceeded the 40% benchmark.

In this setting, individual p-values are **not interpreted as measures of scientific evidence**. Instead, they are treated as **binary indicators of whether a simulated trial meets the predefined success criterion**. Repeating this process across simulations yields an estimate of the **probability of trial success** under each assumed scenario.

4.5.1 Consideration of Multiple Testing and Family-Wise Error Rate

During the design of the simulation study, the issue of multiple testing and family-wise error rate (FWER) was considered. In Exercise 2, the primary analysis is repeated across a large number of simulated trials and across multiple hazard-shape scenarios. At first glance, this repetition could appear analogous to performing multiple hypothesis tests, raising the question of whether corrections such as Bonferroni or Šidák adjustments are required.

However, such corrections are not appropriate in this context. The repeated hypothesis tests conducted in Exercise 2 are not simultaneous tests of different scientific hypotheses on the same dataset. Instead, they represent independent replications of the same pre-specified primary analysis applied to independently simulated datasets. The objective is not to control the probability of making at least one false discovery across tests, but to estimate the probability that a single future trial would meet its success criterion, given the assumed data-generating mechanism.

In other words, each simulated trial corresponds to a hypothetical realization of the same study design, analyzed using the same decision rule. The proportion of simulations in which the null hypothesis is rejected directly estimates the trial's power under the assumed treatment effect. Applying Bonferroni or Šidák corrections in this setting would therefore be conceptually incorrect and would distort the operating characteristics being evaluated.

Accordingly, no multiplicity adjustment is applied in Exercise 2. This approach is standard in simulation-based power and design evaluations, where repeated testing is an intrinsic feature of estimating trial performance rather than a source of inflated type I error.

4.6 Simulation Results

Across all hazard-shape scenarios, the simulated trials consistently reproduced the assumed 24-month survival rate of approximately 60%. While the timing of events differed across scenarios, reflecting differences in hazard shape, the overall survival benefit remained unchanged by construction.

Most importantly, the proportion of simulated trials meeting the success criterion was effectively **100% across all scenarios**. This indicates that the proposed trial design has a very high likelihood of demonstrating benefit if the assumed treatment effect is correct, and that trial success is **robust to uncertainty in the underlying hazard structure**.

4.7 Interpretation and Implications

The results of Exercise 2 show that, conditional on a true 24-month survival rate of 60%, the proposed trial design is highly reliable and insensitive to plausible variations in event timing. The consistently high probability of trial success indicates that **statistical power is not the limiting factor**.

Instead, the primary source of uncertainty lies in the **validity of the assumed treatment effect itself**, rather than in trial mechanics, censoring behavior, or hazard dynamics. From a development perspective, this shifts the key risk from statistical design to clinical assumptions.

4.8 Summary of Exercise 2

Exercise 2 separates belief about treatment efficacy from uncertainty arising due to trial design. By starting with minimal assumptions, rejecting unjustified complexity, and explicitly stress-testing plausible departures from constant hazard behavior, the analysis provides a transparent and decision-relevant assessment of trial robustness.

If the assumed treatment effect holds, the trial design is highly likely to succeed; failure would more likely reflect over-optimism in the assumed effect size rather than deficiencies in statistical design.

5. Integrated Perspective: Exercises 1 and 2 Together

- **Exercise 1** provides empirical evidence that the drug can meaningfully improve survival in lung cancer.
- **Exercise 2** evaluates how reliably a future trial would demonstrate benefit *conditional on a belief* about treatment effect in a different context.

Together, they separate:

- **What the data show**
- **How trial performance depends on assumed treatment effects**

6. Final Conclusions

1. The study drug demonstrates statistically and clinically meaningful improvement in 24-month survival in lung cancer based on observed data.
2. Conditional on a plausible belief about treatment effect, the proposed trial design in a new disease setting is robust and highly likely to succeed.
3. The primary development risk lies not in statistical power or hazard uncertainty, but in the magnitude of the true treatment effect.
4. The strategic application of the same investigational drug across both lung and liver cancers strongly suggests a **histology-independent mechanism of action**, where efficacy is driven by a shared molecular target (such as a specific genetic mutation) or systemic immune activation (immunotherapy) rather than tissue-specific characteristics. This **tumor-agnostic approach** implies that the drug is not only capable of penetrating diverse biological microenvironments—ranging from the air-filled structures of the lung to the dense, vascular tissue of the liver—but also possesses high **metabolic stability**, evidenced by its ability to remain effective within the body's primary detoxification organ. Consequently, success in these distinct indications points toward a robust therapeutic profile with potential utility in treating widespread metastatic disease, regardless of the primary tumor site.

Appendix(R-code):https://drive.google.com/file/d/1LvKYXMhGe02ynG5cqtcll7igOeFoo4Xn/view?usp=sharing