

Article

An Alternative Treatment Effect Measure for Time-to-Event Oncology Randomized Trials

Alan D. Hutson *  and Han Yu

Roswell Park Comprehensive Cancer Center, Department of Biostatistics and Bioinformatics, Elm and Carlton Streets, Buffalo, NY 14623, USA

* Correspondence: alan.hutson@roswellpark.org

Simple Summary

This work introduces a new survival analysis measure called the univariate martingale residual (UMR) for comparing treatments in phase III oncology trials. Traditional methods like hazard ratios and log-rank tests often rely on assumptions such as proportional hazards and large sample sizes, which may not hold in complex trial settings. The UMR provides an unbiased summary of treatment effects by quantifying the average difference between observed and expected events. It also enables exact statistical testing without relying on model-based assumptions. Overall, the UMR offers a more robust and interpretable alternative for evaluating treatment efficacy in challenging survival analysis scenarios.

Abstract

Background/Objectives: Time-to-event endpoints such as Overall Survival (OS), Progression-Free Survival (PFS), and Event-Free Survival (EFS) are central in phase III oncology trials. Hazard ratios from Cox proportional hazards models and log-rank tests are the standard analytic tools, supplemented by Kaplan–Meier estimates. However, these methods depend on proportional hazards to deliver unbiased estimates of treatment effects and large-sample assumptions, and may perform poorly under heavy censoring or non-proportional hazards. We introduce the univariate martingale residual (UMR) as a new endpoint and summary measure that enables exact inference through randomization testing. **Methods:** The UMR reflects the difference between observed and expected events at the subject level. Average UMRs per treatment arm provide an absolute measure of excess events. A randomization-based testing framework is used to compare treatment arms and compute exact *p*-values without proportional hazards or asymptotic assumptions. Performance is assessed through simulations and demonstrated using real oncology trial data. **Results:** UMRs offered robust and interpretable treatment summaries under heavy censoring, non-proportional hazards, and quasi-complete separation, where Cox-based estimates were unstable or undefined. The exact UMR-based randomization test maintained Type I error control and was competitive or more powerful than the log-rank test when proportional hazards were violated. **Conclusions:** The UMR provides an intuitive, assumption-free summary of treatment effects and supports exact inference. It represents a practical and robust alternative to hazard-ratio-based methods in phase III oncology trials, especially in complex survival settings.

Keywords: survival analysis; martingale residuals; randomization test; proportional hazards; permutation test



Academic Editor: Vesna Zadnik

Received: 20 October 2025

Revised: 18 November 2025

Accepted: 20 November 2025

Published: 24 November 2025

Citation: Hutson, A.D.; Yu, H. An Alternative Treatment Effect Measure for Time-to-Event Oncology Randomized Trials. *Cancers* **2025**, *17*, 3750. <https://doi.org/10.3390/cancers17233750>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In phase III oncology trials, time-to-event endpoints are essential for evaluating treatment efficacy. The most common endpoints include Overall Survival (OS), Progression-Free Survival (PFS), and Event-Free Survival (EFS). OS, defined as the time from randomization to death from any cause, is considered the gold standard due to its objectivity, although it typically requires long follow-up and can be influenced by subsequent therapies. PFS measures the time to disease progression or death and is widely used, especially in metastatic settings, though it depends on consistent imaging schedules and standardized progression criteria. EFS, which captures the time to predefined events such as progression, recurrence, or death, is frequently used in early-stage cancers or adjuvant settings, such as in leukemia or lymphoma.

Additional commonly used endpoints include Disease-Free Survival (DFS), which focuses on recurrence or death after curative treatment; Time to Progression (TTP), which measures progression while excluding non-cancer-related deaths; Time to Treatment Failure (TTF), which captures treatment discontinuation due to progression, toxicity, or other reasons; and Duration of Response (DoR), which quantifies the duration of tumor response in patients who initially respond to treatment.

Endpoint selection depends on cancer type, disease stage, trial objectives, and regulatory requirements. OS and PFS are most commonly prioritized due to their clinical relevance and regulatory acceptance, while EFS and DFS are prevalent in early-stage or adjuvant trials. TTP, TTF, and DoR typically serve as secondary endpoints.

In this note, we focus on straightforward two-group comparisons in randomized phase III trials. Without loss of generality to more complex designs, we introduce a new endpoint termed the univariate martingale residual for comparing survival curves that facilitates both nonparametric estimation of treatment effects per arm and exact Type I error control within the framework of a randomization test. The methods discussed are also applicable to designs that incorporate stratification or to earlier-phase trials.

For time-to-event endpoints in two-group comparisons, the primary statistical measure of treatment effect is the hazard ratio. This is typically estimated using the Cox proportional hazards model and quantifies the relative risk of an event between treatment arms under the proportional hazards assumption. These estimates are usually accompanied by 95% confidence intervals to reflect estimation precision. The log-rank test is commonly used to compare survival curves and assess the statistical significance of differences in event times over the follow-up period. In the two-group setting, the log-rank test and the score test from the Cox model yield equivalent *p*-values under the null hypothesis of equal hazards [1].

Median survival times are derived from Kaplan–Meier (KM) estimates and indicate the time by which 50% of participants have experienced the event. These are also reported with 95% confidence intervals. Event-free probabilities at fixed time points, also derived from KM curves, provide clinically interpretable measures of the proportion of patients who remain event-free at specified intervals.

The Cox model and the log-rank test for comparing treatment arms both rely on the proportional hazards assumption. While this assumption does not affect Type I error control in randomized trials under the null hypothesis, provided that asymptotic approximations are accurate, it can lead to substantial power loss under alternatives. In particular, when hazard rates over time between the two groups are not proportional, the hazard ratio may provide a biased and potentially misleading summary of treatment effect at any specific time point. Although these methods are powerful when the assumptions hold, they depend heavily on large-sample approximations.

Several studies have reported inflated Type I error rates associated with the Cox model in small-sample settings [2–4]. As Shao et al. [4] observe, the Cox model often performs

poorly when the number of events is small or when there is heavy censoring in one or both treatment groups. In response to these limitations, Shao et al. proposed an exact test based on the Cox model. However, this test still assumes proportional hazards to achieve optimal power and to support unbiased estimation of the hazard ratio.

Alternative point estimates, such as median survival time, may be undefined in the presence of censoring. Even when defined, they may fail to capture important differences in survival if the curves overlap at the median despite diverging elsewhere. Similarly, survival probabilities at fixed time points may miss broader differences between groups if the curves cross at those specific times.

In order to get a sense of what is being used in phase III oncology trials we examined a series of 30 phase III oncology trials published sequentially in the *New England Journal of Medicine* from 2023–2025 [5–34]. Across the 30 phase III clinical trials, survival rates at specific time points were reported as follows: no trials (0%) reported 6-month survival rates, 2 trials (6.6%) reported 1-year rates for progression-free survival (PFS) or treatment failure-free survival, 1 trial (3.3%) reported 18-month event-free survival (EFS), and 7 trials (23.3%) reported 2-year rates for PFS, EFS, disease-free survival (DFS), or overall survival (OS), making it the most common time point. Additionally, 2 trials (6.6%) reported 3-year rates for PFS, OS, or invasive disease-free survival (IDFS), 1 trial (3.3%) reported 5-year DFS, 1 trial (3.3%) reported 8-year OS, and 1 trial (3.3%) reported 10-year OS and local recurrence rates. All survival probabilities and median survival times (e.g., median PFS, OS, EFS, or DFS) were calculated using the Kaplan–Meier method, with hazard ratios (HRs) estimated by stratified Cox proportional-hazards models in all 30 trials (100%) to compare treatment groups, often stratified by factors like mutation status or disease stage. The log-rank test was used in 29 trials (96.7%) to compare survival distributions, often stratified, with two trials focusing on noninferiority analyses. The 2-year rates were most frequently reported due to their relevance in assessing treatment efficacy in advanced or metastatic cancers, while longer-term rates (5–10 years) were less common, reflecting shorter median follow-up periods (7.8–62 months) in most trials.

In this note, we introduce and evaluate a new metric, termed the **univariate martingale residual** (UMR) [35], considering both its use as a test statistic and as a summary measure of treatment effect within each treatment arm. In essence, the martingale residual quantifies the difference between the observed and expected number of events. Subject-level UMRs can be grouped by treatment arm in a randomized study under the null hypothesis of no treatment effect, and comparisons may be made by examining the average UMRs across arms. Further details are provided in the next section.

In this note, we outline the development and application of the UMR. Section 2 introduces its formal definition. Section 3 illustrates the randomization test based on UMRs in a two-group setting. In Section 4, we present simulated examples and a corresponding power study. Section 5 provides three real-world applications, followed by concluding remarks in Section 6.

2. Definition of Univariate Martingale Residuals

We briefly review a recently proposed randomization test framework for right-censored outcomes, originally developed by Hutson and Yu [35], which integrates a martingale-residual-based approach into the classical randomization test setting. This method allows for exact Type I error control and is especially advantageous when standard assumptions such as proportional hazards are violated or when the sample size is small.

In the simple parallel-group design, the randomization test proceeds by considering the outcome variable of interest, in this case, the univariate martingale residual (UMR) defined below, under the strict additivity framework of Kempthorne [36]. Suppose

$N = n + m$ subjects are randomized to treatments A and B , respectively, with all $\binom{N}{n}$ assignments equally likely. Let $t_1, t_2, \dots, t_N$ represent independent and identically distributed failure times, and let $c_1, c_2, \dots, c_N$ denote the corresponding non-informative right censoring times, independently distributed and indexed by $i = 1, 2, \dots, N$, pooled across treatments A and B . Due to right censoring, we only observe $g \leq N$ of the t 's, where $y_i = \min(t_i, c_i)$ and $\delta_i = I_{(t_i > c_i)}$. The UMR for subject $i, i = 1, 2, \dots, N$, is defined as:

$$r_i = \delta_i - \hat{H}(y_i), \quad (1)$$

where δ_i is the event indicator and $\hat{H}(y_i) = -\log(\hat{S}(y_i))$ is the estimated cumulative hazard at time t_i , obtained from the product-limit estimator of the survival function. This estimator, derived from discretized hazard functions, follows the classic form:

$$\hat{S}(y) = \prod_{y_{(j)} \leq y} (1 - \hat{h}_j), \quad (2)$$

with $\hat{h}_j = d_j / r_j$, where d_j and r_j are the number of events and number at risk at time $y_{(j)}$, respectively [37].

Martingale residuals, originally discussed in the context of model diagnostics for the Cox regression model [38], have the appealing property that $\sum_{i=1}^N r_i = 0$. In the randomization test framework, these residuals can be treated as fixed outcomes [39], with randomness only in the treatment assignment. In the univariate setting the UMR's also exhibit the property $\sum_{i=1}^N r_i = 0$ [38].

The UMR, as defined at (1), can be interpreted at each observed y_i as the expected number of excess deaths over $[0, t]$. When the UMRs are delineated by treatment arm we would expect a poorly performing treatment to have a large proportion of observed r_i 's greater than zero and a efficacious treatment to have a large proportion of observed r_i 's less than zero given the pooled sample. The observed UMR values form the basis for constructing an exact test comparing treatment arms. An advantage of using UMRs, as noted in Hutson and Yu [35], is that they remain estimable even when median survival times are undefined due to heavy censoring. Their mean and standard deviation can be directly computed by treatment arm, offering interpretable summary statistics that retain efficiency and robustness across a variety of censoring scenarios. UMRs can be easily output using standard statistical software packages for fitting a Cox regression model by simply using a dummy variable covariate set to one and outputting the martingale residuals in a standard fashion.

3. Randomization Testing with UMRs

The null hypothesis for comparing treatment arms A and B using UMRs can be written as

$$H_0 : F_A(s) = F_B(s), \quad \forall s,$$

where $F_A(s)$ and $F_B(s)$ are the empirical cumulative distribution functions (c.d.f.s) of the UMRs given fixed observations for treatments A and B , respectively. Standard test statistics such as the Wilcoxon rank-sum or a randomization t -statistic may be used. In this work, we employ the difference in average Martingale residuals between groups A and B .

In the parallel-group setting, the randomization test proceeds as follows. For any outcome variable, discrete, ordinal, or continuous, an exact randomization test is applicable. Here, the outcome variable is the UMR. Let $N = n + m$, where n and m denote the numbers assigned to treatments A and B , respectively. There are $\binom{N}{n}$ equally likely assignments, each with probability $1/\binom{N}{n}$.

Let $I_i \in \{0, 1\}$ denote the treatment indicator for subject i ($i = 1, \dots, N$), with $P(I_i = 1) = \pi \in (0, 1)$. Following Rosenbaum [39], the observed UMR is

$$R_i = I_i r_{A_i} + (1 - I_i) r_{B_i},$$

where $I_i = 1$ indicates treatment A and $I_i = 0$ indicates treatment B . Under strict additivity [36], the $\{r_i\}$ are fixed, with randomness arising solely from the $\{I_i\}$.

We define $r_{A_i} = s_i + a_i$ and $r_{B_i} = s_i + b_i$, where s_i is the subject effect and a_i, b_i the treatment effects. Thus, if both treatments were hypothetically applied to subject i , the difference $r_{A_i} - r_{B_i} = a_i - b_i$ would be constant across subjects. Inference is therefore driven by randomization.

For one-sided testing,

$$H_0 : F_A(s) = F_B(s), \quad \forall s \quad \text{vs.} \quad H_1 : F_A(s) > F_B(s), \quad \forall s,$$

(or analogously $H_1 : F_A(s) < F_B(s)$). Under a shift alternative, where both c.d.f.s belong to the same family, this is equivalent to testing about the expectations and expressed as

$$H_0 : \mathbb{E}[R \mid I = 1] = \mathbb{E}[R \mid I = 0] \quad \text{vs.} \quad H_1 : \mathbb{E}[R \mid I = 1] > \mathbb{E}[R \mid I = 0].$$

Define the observed test statistic

$$S = \bar{r}_A - \bar{r}_B, \quad \bar{r}_A = \frac{1}{n} \sum_{j=1}^N I_j R_j, \quad \bar{r}_B = \frac{1}{m} \sum_{j=1}^N (1 - I_j) R_j.$$

For each of the $l = \binom{N}{n}$ permutations, compute

$$S_k = \bar{r}_{A,k} - \bar{r}_{B,k}, \quad k = 1, \dots, l.$$

The exact p -values are

$$p = \frac{1}{l} \sum_{k=1}^l \mathbf{1}\{S_k \geq S\}, \quad \text{for } H_1 : \mathbb{E}[R \mid I = 1] > \mathbb{E}[R \mid I = 0],$$

$$p = \frac{1}{l} \sum_{k=1}^l \mathbf{1}\{S_k \leq S\}, \quad \text{for } H_1 : \mathbb{E}[R \mid I = 1] < \mathbb{E}[R \mid I = 0],$$

where $\mathbf{1}\{\cdot\}$ is the indicator function.

For a two-sided alternative,

$$H_1 : \mathbb{E}[R \mid I = 1] \neq \mathbb{E}[R \mid I = 0],$$

a natural test statistic is

$$S = |\bar{r}_A - \bar{r}_B|,$$

with p -value

$$p = \frac{1}{l} \sum_{k=1}^l \mathbf{1}\{S_k^2 \geq S^2\}.$$

In practice, p -values are often estimated via B Monte Carlo permutations, replacing l with B . Typically $B = 10,000$ yields stable estimates.

Type I Error Control. Each configuration of the UMR vector $\mathbf{r}$ has probability $1/\binom{N}{n}$. Let S^* denote the test statistic under a re-randomization. We test

$$H_0 : F_A(s) = F_B(s), \quad \forall s$$

against either

$$H_1 : F_A(s) > F_B(s), \quad \forall s, \quad \text{or in the location-shift model} \quad H_1 : \mathbb{E}[R \mid I = 1] > \mathbb{E}[R \mid I = 0].$$

The critical value t_α is chosen such that

$$\Pr_{H_0}(S^* \geq t_\alpha) \leq \alpha.$$

Let s_α be the threshold yielding a significance level as close to α as possible. Without ties, the exact level is

$$\frac{\lfloor \alpha \binom{N}{n} \rfloor}{\binom{N}{n}},$$

where $\lfloor \cdot \rfloor$ is the integer part. With ties near t_α , the level may be slightly lower.

Comment on Blocking and Stratification. If blocked randomization or stratification is used, the test is modified by permuting treatment assignments within blocks and/or strata.

4. Simulation Study

4.1. Examples

To illustrate the proposed methods by mimicking realistic settings, we consider several simulation scenarios based on Weibull-distributed random variables. Let Y_B and Y_A denote random variables with quantile functions

$$Q_B(u) = [-\sigma_B \log(1 - u)]^{\theta_B}, \quad Q_A(u) = [-\sigma_A \log(1 - u)]^{\theta_A},$$

where $\sigma_B, \sigma_A > 0$ and $\theta_B, \theta_A > 0$. All tests used in the examples were two-sided.

Hazard Functions

For a Weibull quantile function $Q(u) = [-\sigma \log(1 - u)]^\theta$, the corresponding hazard function is

$$h(y) = \frac{1}{\theta\sigma} y^{\frac{1-\theta}{\theta}}, \quad y > 0.$$

Thus:

- For Y_B : $h_B(y) = \frac{1}{\theta_B\sigma_B} y^{\frac{1-\theta_B}{\theta_B}}$,
- For Y_A : $h_A(y) = \frac{1}{\theta_A\sigma_A} y^{\frac{1-\theta_A}{\theta_A}}$.

Hazard Ratio

The hazard ratio of Y_B relative to Y_A is

$$\frac{h_B(y)}{h_A(y)} = \frac{\theta_A\sigma_A}{\theta_B\sigma_B} y^{\frac{1-\theta_B}{\theta_B} - \frac{1-\theta_A}{\theta_A}}.$$

This ratio is constant only when $\theta_B = \theta_A$, in which case it simplifies to $\frac{\sigma_A}{\sigma_B}$.

Censoring is introduced via an exponential random variable C with quantile function

$$Q_C(u) = -\sigma_C \log(1 - u).$$

Scenario 1

We consider $n = m = 75$ with parameters:

$$\sigma_0 = 2, \quad \theta_0 = 1, \quad \sigma_1 = 1, \quad \theta_1 = 1, \quad \sigma_C = 1.$$

Results.

- **Group means of martingale residuals:** $A = -0.142$, $B = 0.142$.
- **Randomization test (UMR difference):** $p = 0.0104$.
- **Median survival:** 1.096 for A, 0.615 for B.
- **Log-rank test:** $\chi^2 = 6.389$, $p = 0.0115$.
- **Cox PH model:** $\widehat{HR} = 1.809$ with 95% CI (1.135, 2.882), $p = 0.0127$.

The Kaplan–Meier curves are shown in Figure 1. Under proportional hazards, the three tests yield consistent inference. The UMRs suggest that Arm A experienced an average of -0.142 excess deaths over $[0, t]$ while Arm B experienced an average of 0.142 , providing a clear interpretation of treatment differences. Inference across the log-rank, Cox Model and randomization test provide similar results.

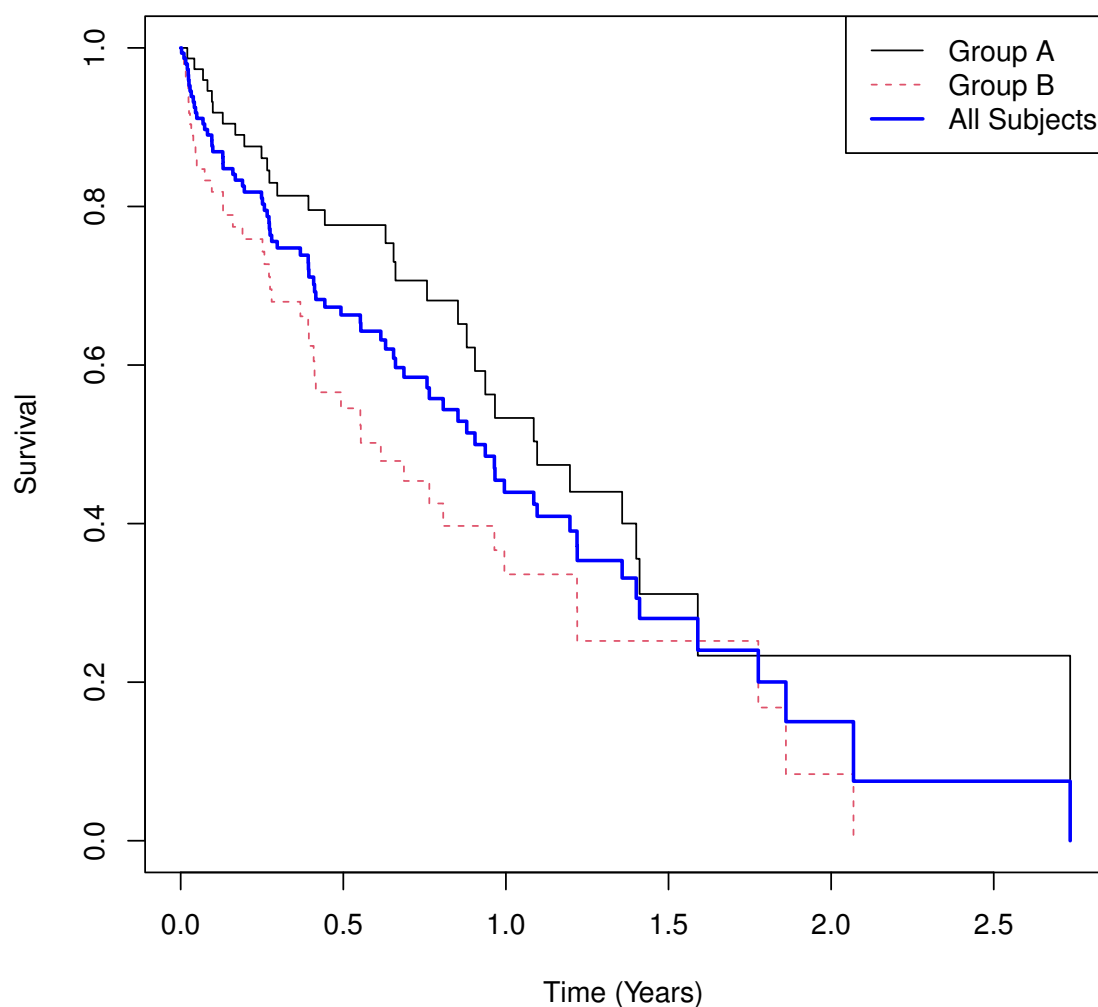


Figure 1. Simulated Example 1: $n = m = 75$, $\sigma_A = 2$, $\theta_A = 1$, $\sigma_B = 1$, $\theta_B = 1$, and $\sigma_C = 1$.

Scenario 2

We consider $n = m = 75$ with parameters:

$$\sigma_0 = 2, \quad \theta_0 = \frac{1}{2}, \quad \sigma_1 = 1, \quad \theta_1 = 1, \quad \sigma_C = 1.$$

Results.

- **Group means of martingale residuals:** $A = -0.117$, $B = 0.117$.
- **Randomization test (UMR difference):** $p = 0.0285$.
- **Median survival:** 1.121 for A, 0.977 for B.
- **Log-rank test:** $\chi^2 = 4.780$, $p = 0.0288$.
- **Cox PH model:** $\widehat{HR} = 1.725$ with 95% CI (1.052, 2.829), $p = 0.0307$.

One-year survival probabilities are estimated as $\hat{S}_A(1) = 0.54$ and $\hat{S}_B(1) = 0.48$. The Kaplan–Meier curves (Figure 2) reveal clear non-proportional hazards as expected based on the simulation parameters. Median survival and fixed-time probabilities provide limited insight, while the UMRs again yield a simple and interpretable summary.

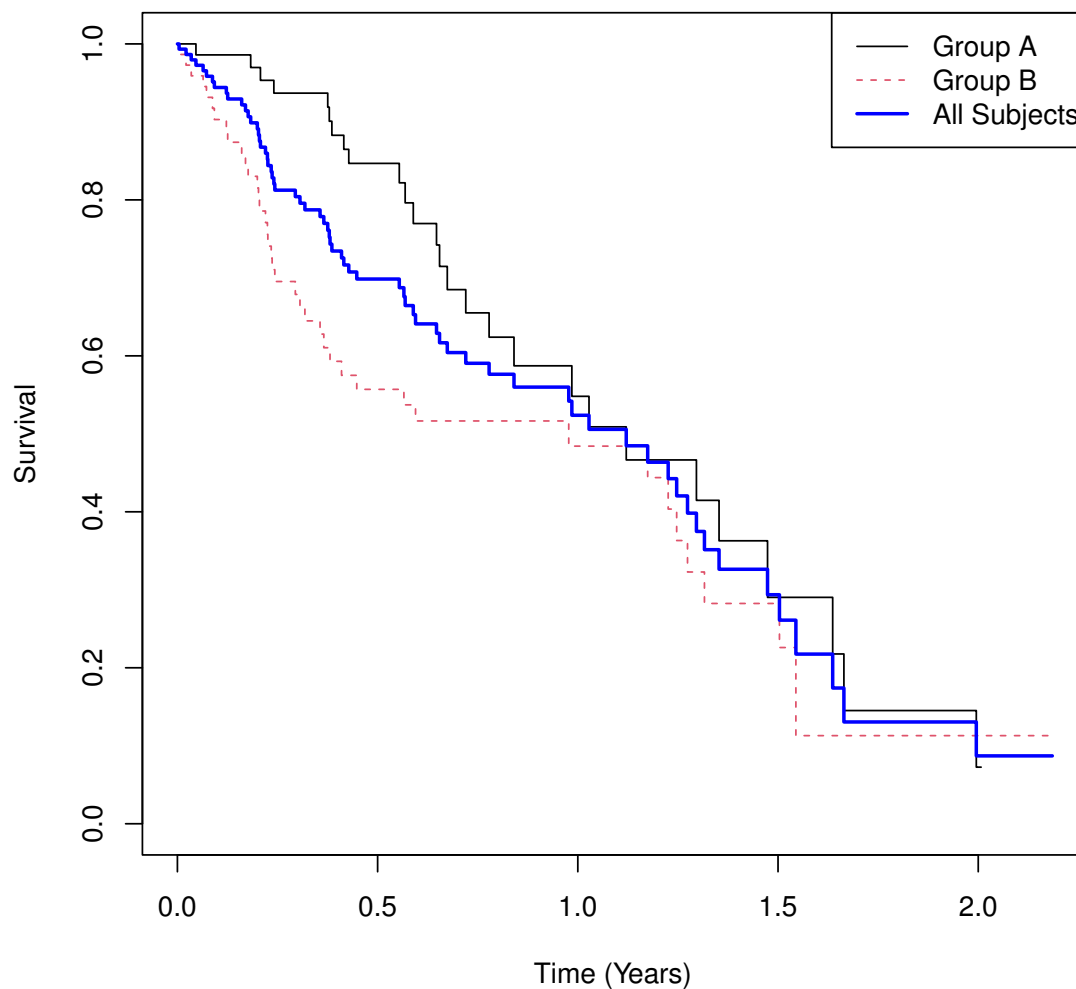


Figure 2. Simulated Example 2: $n = m = 75$, $\sigma_A = 2$, $\theta_A = 1/2$, $\sigma_B = 1$, $\theta_B = 1$, and $\sigma_C = 1$.

Scenario 3

We consider $n = m = 75$ with parameters:

$$\sigma_0 = 3, \quad \theta_0 = \frac{1}{2}, \quad \sigma_1 = 1, \quad \theta_1 = 1, \quad \sigma_C = \frac{1}{4}.$$

Results.

- **Group means of martingale residuals:** $A = -0.089$, $B = 0.089$.
- **Randomization test (UMR difference):** $p = 0.0007$.
- **Median survival:** not estimable for either group due to heavy censoring.
- **Log-rank test:** $\chi^2 = 10.783$, $p = 0.0010$.
- **Cox PH model:** $\widehat{HR} = 8.191$ with 95% CI (1.853, 36.208), $p = 0.0055$.

The Kaplan–Meier curves (Figure 3) confirm that median survival and fixed-time survival probabilities are uninformative. Non-proportional hazards lead to a biased HR estimate, while UMRs provide an interpretable and stable summary: -0.089 excess deaths for A versus 0.089 for B.

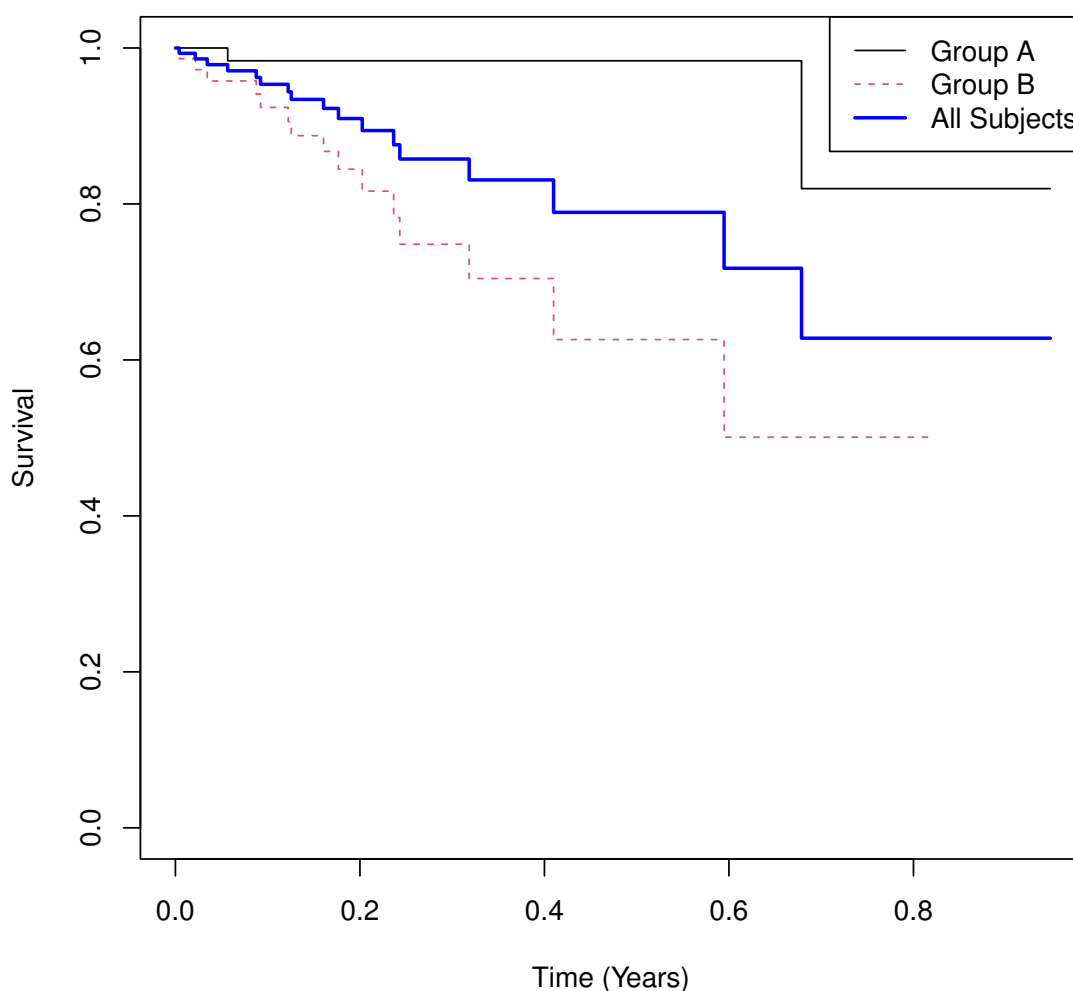


Figure 3. Simulated Example 3: $n = m = 75$, $\sigma_A = 3$, $\theta_A = 1/2$, $\sigma_B = 1$, $\theta_B = 1$, and $\sigma_C = 1/4$.

Scenario 4

We consider $n = m = 75$ with parameters:

$$\sigma_0 = 3000, \quad \theta_0 = 1, \quad \sigma_1 = 1, \quad \theta_1 = 1, \quad \sigma_C = 1.$$

Results.

- **Group means of martingale residuals:** $A = -0.298$, $B = 0.298$.
- **Randomization test (UMR difference):** $p < 0.0001$.
- **Median survival:** not estimable for A, 0.977 for B.
- **Log-rank test:** $\chi^2 = 53.562$, $p < 0.0001$.
- **Cox PH model:** $\widehat{HR} \approx 9.08 \times 10^8$ with 95% CI $(0, \infty)$, $p = 0.996$.

Here, no events occur in Arm A, a situation often encountered in practice. The Kaplan–Meier curves (Figure 4) show that the HR is not estimable, and the Cox model p -value is unreliable due to quasi-separation. In contrast, UMRs again yield a stable and interpretable measure: -0.298 excess deaths for A versus 0.298 for B, with inference consistent with the log-rank test.

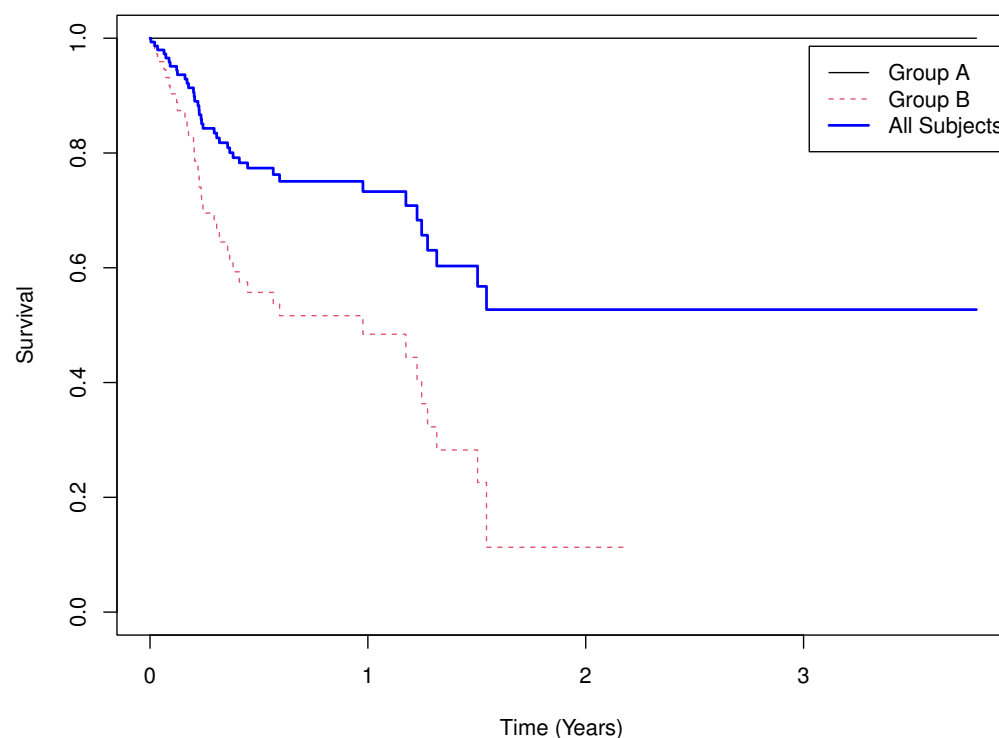


Figure 4. Simulated Example 4: $n = m = 75$, $\sigma_A = 3000$, $\theta_A = 1$, $\sigma_B = 1$, $\theta_B = 1$, and $\sigma_C = 1$.

Across all scenarios, UMRs provide a consistent summary measure and enable exact randomization inference, even when proportional hazards are violated or Cox models suffer from numerical instability. The next section evaluates type I error control and power.

4.2. Power Comparison

To assess the operating characteristics of the proposed randomization test, we conducted a simulation study under a variety of parameter settings. In each replicate, survival times were generated from Weibull-type models with varying scale parameters (σ_0 , σ_c) and shape parameter θ_0 , while the comparison group was fixed at $\sigma_1 = 1$ and $\theta_1 = 1$. Each simulated dataset consisted of $n = m = 100$ subjects per group, and censoring times were drawn to reflect different censoring distributions. For each scenario, $N = 1000$ replicates were performed, and the randomization test was carried out with 1000 resamples. Both the log-rank test and the randomization test based on differences in average martingale residuals were evaluated at a significance level of $\alpha = 0.05$ (two-sided).

In general, the simulation results show that the log-rank test and the randomization test produce nearly identical rejection probabilities across all parameter combinations

considered. This suggests that, in practice, the two methods perform equivalently, and thus an extensive simulation study is unnecessary beyond illustrative examples. As seen in Tables 1 and 2, the estimated powers of the two procedures are consistently aligned, confirming that the permutation approach retains the operating characteristics of the traditional log-rank test while offering a useful alternative formulation. From a heuristic perspective, these findings further suggest that average UMRs may serve as an appropriate summary measure to pair with the log-rank test, or that the randomization test alone may be employed as a practical substitute.

Table 1. Simulation rejection proportions for permutation and logrank tests under varying $(\sigma_0, \sigma_c, \theta_0)$: Part I.

σ_0	σ_c	θ_0	Permutation	Logrank
0.5	0.5	0.5	0.055	0.058
1.0	0.5	0.5	0.441	0.449
1.5	0.5	0.5	0.827	0.827
2.0	0.5	0.5	0.949	0.952
0.5	1.0	0.5	0.344	0.369
1.0	1.0	0.5	0.148	0.148
1.5	1.0	0.5	0.615	0.621
2.0	1.0	0.5	0.895	0.900
0.5	1.5	0.5	0.614	0.654
1.0	1.5	0.5	0.052	0.053
1.5	1.5	0.5	0.412	0.420
2.0	1.5	0.5	0.796	0.800
0.5	0.5	1.0	0.888	0.888
1.0	0.5	1.0	0.048	0.049
1.5	0.5	1.0	0.340	0.345
2.0	0.5	1.0	0.680	0.692
0.5	1.0	1.0	0.958	0.959
1.0	1.0	1.0	0.060	0.063
1.5	1.0	1.0	0.443	0.443
2.0	1.0	1.0	0.882	0.884
0.5	1.5	1.0	0.982	0.982
1.0	1.5	1.0	0.048	0.047
1.5	1.5	1.0	0.567	0.566
2.0	1.5	1.0	0.934	0.941

Table 2. Simulation rejection proportions for permutation and logrank tests under varying $(\sigma_0, \sigma_c, \theta_0)$: Part II.

σ_0	σ_c	θ_0	Permutation	Logrank
0.5	0.5	1.5	0.996	0.996
1.0	0.5	1.5	0.215	0.221
1.5	0.5	1.5	0.100	0.102
2.0	0.5	1.5	0.352	0.355
0.5	1.0	1.5	0.999	0.999
1.0	1.0	1.5	0.111	0.109
1.5	1.0	1.5	0.278	0.277
2.0	1.0	1.5	0.749	0.749
0.5	1.5	1.5	0.997	0.997
1.0	1.5	1.5	0.054	0.057
1.5	1.5	1.5	0.472	0.484
2.0	1.5	1.5	0.931	0.936
0.5	0.5	2.0	1.000	1.000
1.0	0.5	2.0	0.526	0.526
1.5	0.5	2.0	0.042	0.047
2.0	0.5	2.0	0.209	0.215
0.5	1.0	2.0	1.000	1.000
1.0	1.0	2.0	0.237	0.241
1.5	1.0	2.0	0.176	0.181
2.0	1.0	2.0	0.658	0.665
0.5	1.5	2.0	1.000	1.000
1.0	1.5	2.0	0.108	0.110
1.5	1.5	2.0	0.401	0.412
2.0	1.5	2.0	0.887	0.893

5. Real World Examples

5.1. Example 1

Our first real-world example is based on a randomized trial comparing androgen-deprivation therapy (ADT) plus docetaxel versus ADT alone in patients with metastatic hormone-sensitive prostate cancer [40]. In this large study of 790 patients, the primary objective was to test the hypothesis that the median overall survival would be 33.3% longer among those receiving early docetaxel in addition to ADT compared with ADT alone. The sample sizes were $n = 397$ for the ADT + docetaxel group ($I = 1$) and $m = 393$ for the ADT-alone group ($I = 0$).

For illustration purposes, the original stratification factors were not used in our analysis. The results are summarized below, and the Kaplan–Meier survival curves for each treatment arm and combined are shown in Figure 5. The permutation UMR test was significant at $\alpha = 0.05$ (two-sided), indicating that patients receiving ADT alone experienced an average expected excess of 0.067 deaths over the study period, while those receiving ADT + docetaxel experienced an expected average decrease of -0.055 deaths. The inference from the log-rank test was consistent with the permutation UMR result. Visual inspection of the survival curves suggests non-proportional hazards, as the curves overlap during the early time period, indicating that the hazard ratio may not provide a fully adequate summary of treatment efficacy.

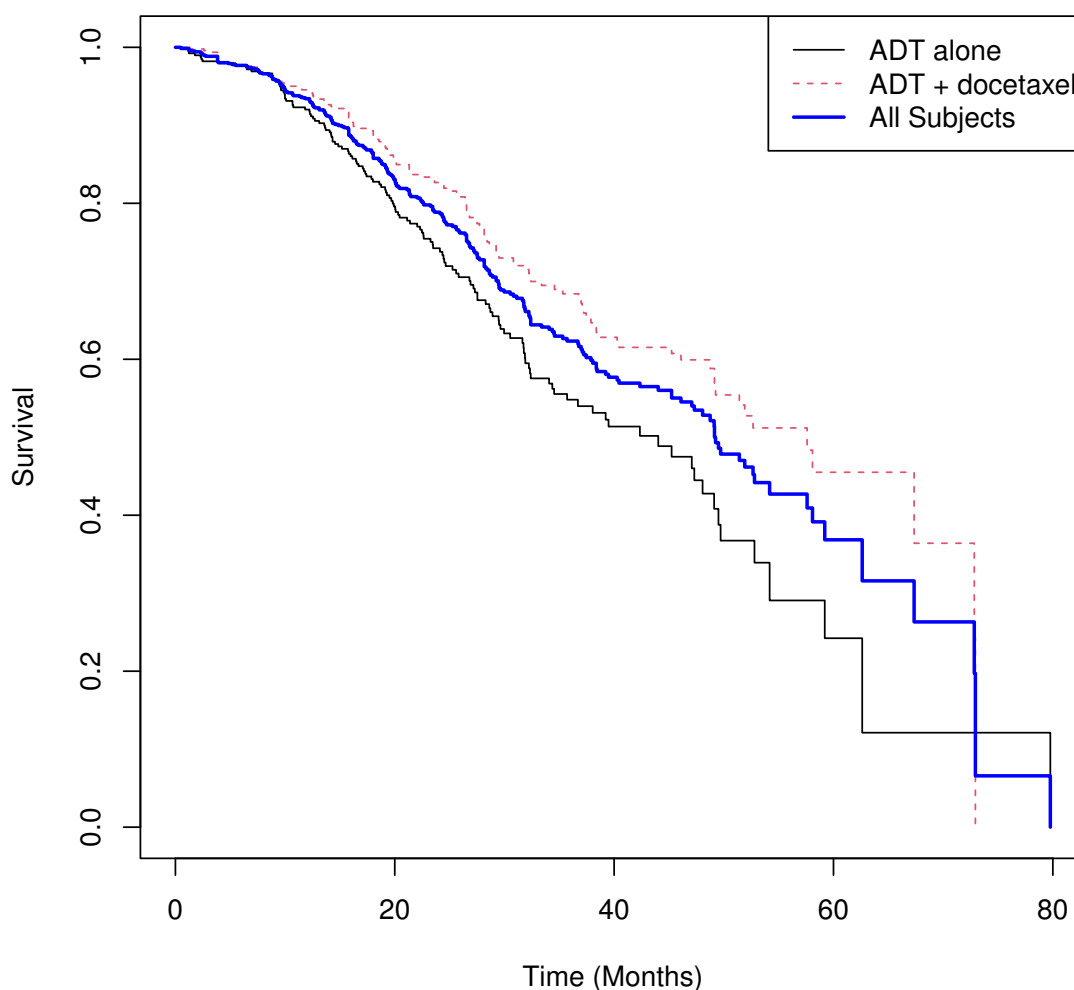


Figure 5. Real World Example 1: Estimated survival curves of ADT plus docetaxel versus those receiving ADT and combined survival curve estimate.

Results.

- **Group means of martingale residuals:** $I = 0$: 0.067, $I = 1$: -0.055 .
- **Randomization test (UMR difference):** two-sided $p = 0.0006$.
- **Median survival:** 44.0 for $z = 0$, 57.6 for $z = 1$.
- **Log-rank test:** $\chi^2 = 11.31$, $p = 0.00077$.
- **Cox PH model:** $\widehat{HR} = 0.664$ with 95% CI (0.520, 0.846), $p = 0.00095$.

5.2. Example 2

For our second example, we utilized publicly available data from Project Data Sphere (<https://data.projectdatasphere.org/>) accessed on July 17, 2020. Specifically, we examined a randomized phase II clinical trial [41] that evaluated the efficacy and safety of LY2510924 (LY) in combination with first-line standard-of-care (SOC) chemotherapy for patients with extensive-disease small cell lung cancer (ED-SCLC). The primary efficacy endpoint of the original study was progression-free survival (PFS), with overall survival (OS) designated as a key secondary endpoint. For this illustration, we focused on OS.

The treatment groups were defined as follows: LY + SOC (Arm A; $n = 47$, $I = 1$) and SOC alone (Arm B; $m = 42$, $I = 0$). The results are summarized below, and the Kaplan–Meier survival curves for each treatment arm and the combined data are presented in Figure 6. The permutation UMR test was not significant at $\alpha = 0.05$ (two-sided). Descriptively, patients receiving LY + SOC experienced an average expected excess of 0.130 deaths over the study period, whereas those receiving SOC experienced an expected an average decrease of -0.146 deaths. The inference from the log-rank test was consistent with the permutation UMR test.

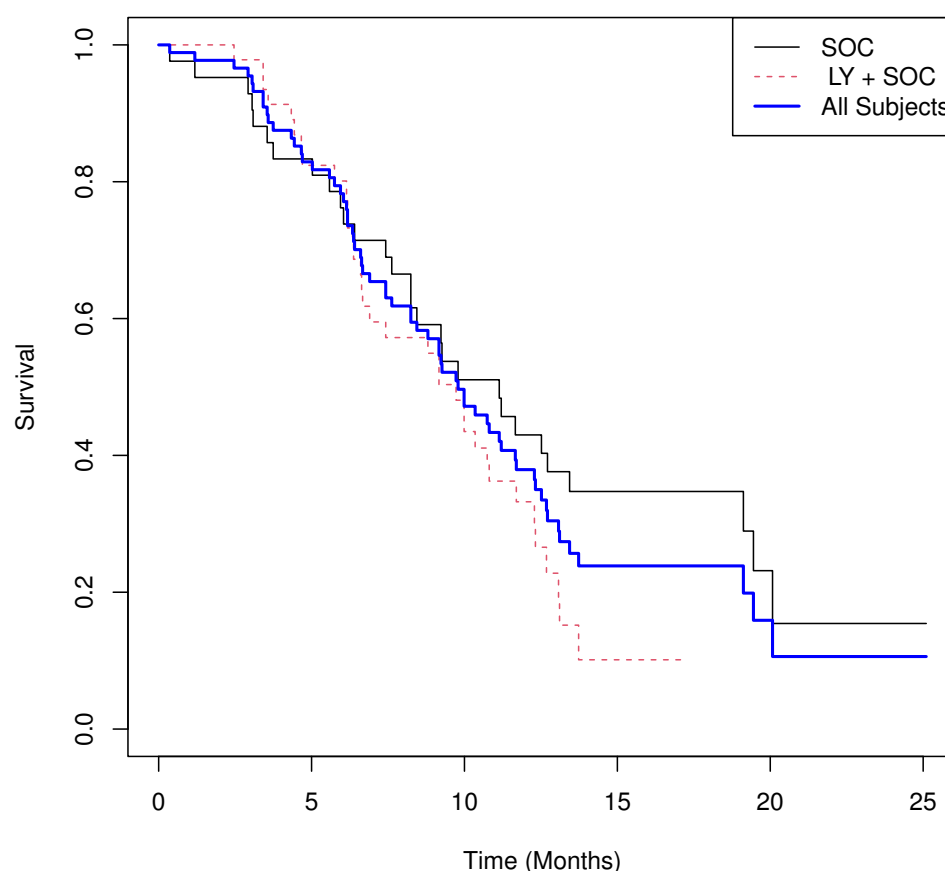


Figure 6. Real World Example 2: Estimated survival curves of LY + SOC versus those receiving SOC and combined survival curve estimate.

Visual inspection of the survival curves suggests clear evidence of non-proportional hazards, as the curves crisscross over time, indicating that the hazard ratio may not serve as an appropriate summary of treatment effect. Similarly, the estimated survival probabilities at six months were comparable between groups: $\hat{S}_0(6) = 0.762$ and $\hat{S}_1(6) = 0.801$.

Results.

- **Group means of martingale residuals:** $I = 0$: -0.146 , $I = 1$: 0.130 .
- **Randomization test (UMR difference):** two-sided $p = 0.121$.
- **Median survival:** 11.1 for $z = 0$, 9.7 for $z = 1$.
- **Log-rank test:** $\chi^2 = 2.53$, $p = 0.112$.
- **Cox PH model:** $\widehat{HR} = 1.52$ with 95% CI (0.905, 2.540), $p = 0.114$.

5.3. Example 3

Our third example is based on a randomized clinical trial involving post-surgical patients with Stage I–IIA cervical cancer who exhibited pathologically confirmed intermediate-risk factors, including combinations of capillary lymphatic space (CLS) involvement, stromal invasion, and tumor size [42]. Patients were randomly assigned in a 1:1 ratio to receive either adjuvant chemoradiotherapy (CRT; $n = 158$) or radiotherapy (RT; $m = 158$). All patients received conformal RT or intensity-modulated radiation therapy (IMRT). In the CRT arm, cisplatin was administered at a dose of 40 mg/m² weekly for six cycles during RT. Recurrence-free survival (RFS) was the primary endpoint among randomized and eligible patients.

Because the raw data were not publicly available, we reconstructed a synthetic dataset with equivalent sample size and censoring distribution using the published Kaplan–Meier curves and the WebPlotDigitizer tool v5 (<https://automeris.io/wpd/>). These reconstructed data are used solely for illustrative purposes. The treatment groups were defined as CRT ($I = 0$) and RT ($I = 1$).

The results are summarized below, and the Kaplan–Meier survival curves for RFS by treatment arm and combined data are shown in Figure 7. As illustrated in Figure 7, the median RFS was not estimable for either treatment arm due to the degree of censoring. The permutation UMR test was not significant at $\alpha = 0.05$ (two-sided). Descriptively, patients receiving CRT experienced an average expected excess of 0.04 deaths over the study period, whereas those receiving RT experienced an average expected decrease of -0.04 deaths. The inference from the log-rank test was consistent with the permutation UMR result.

Results.

- **Group means of martingale residuals:** $I = 0$: 0.040 , $I = 1$: -0.040 .
- **Randomization test (mean difference):** two-sided $p = 0.095$.
- **Median survival:** not estimable for either group.
- **Log-rank test:** $\chi^2 = 2.82$, $p = 0.093$.
- **Cox PH model:** $\widehat{HR} = 0.64$ with 95% CI (0.378, 1.082), $p = 0.096$.

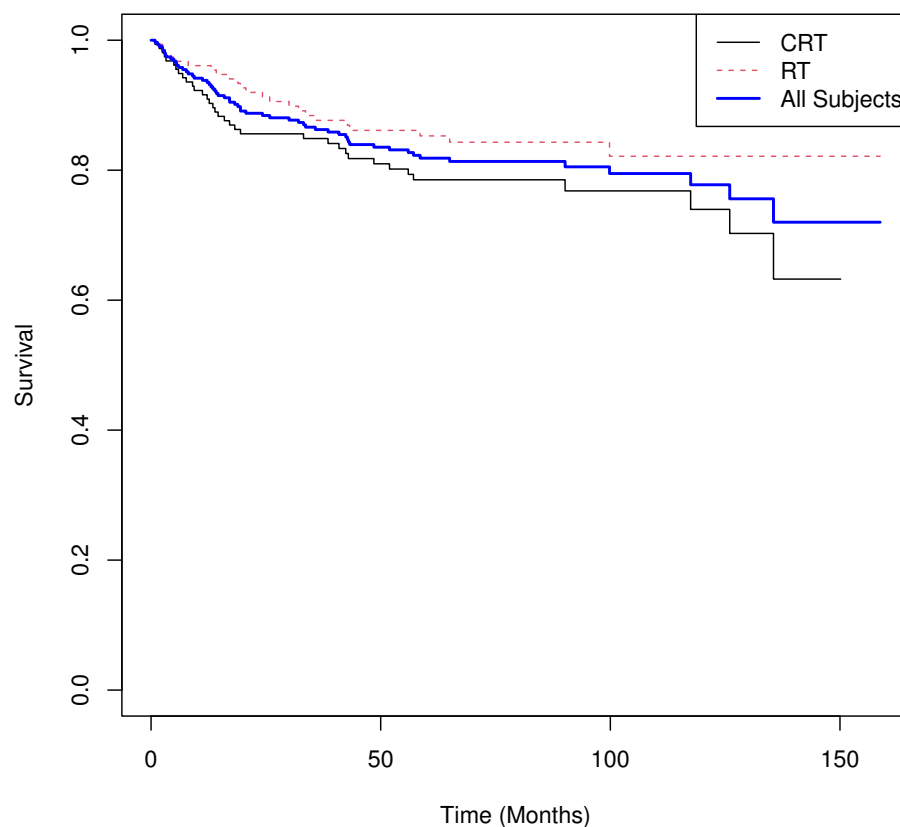


Figure 7. Real World Example 3: Estimated recurrence-free survival of CRT versus those receiving gRT and combined recurrence-free survival curve estimate.

6. Conclusions

This paper introduces a direct and interpretable approach for quantifying treatment effects in survival analysis. The proposed measure, the *average univariate martingale residual* (UMR) per treatment arm, serves as a simple yet powerful summary statistic that represents the average number of excess deaths attributable to treatment differences. Unlike the conventional hazard ratio, which depends on the proportional hazards assumption, the UMR provides an assumption-free, unbiased estimate of treatment efficacy over the entire time course that can be interpreted on an absolute scale.

The method demonstrates robustness in complex and challenging trial settings. In particular, it yields stable and meaningful summaries under heavy censoring, where median survival cannot be reliably estimated, and in cases of quasi-complete separation, where the Cox model and associated hazard ratio estimates become unstable or undefined. These properties make the approach particularly valuable for modern oncology trials and other studies where such data complications are common.

In addition to introducing a new summary metric, we integrate the UMR within an exact randomization (permutation) testing framework to enable valid statistical inference. This combination allows the computation of exact *p*-values without reliance on large-sample approximations or model-based assumptions, in contrast to standard methods such as the log-rank test or Cox model. Collectively, this framework provides a rigorous and interpretable alternative for evaluating treatment effects in randomized clinical trials.

Author Contributions: Conceptualization, A.D.H. and H.Y.; Methodology, A.D.H.; Software, H.Y.; Formal analysis, A.D.H.; Investigation, A.D.H.; Resources, A.D.H.; Writing—original draft, A.D.H.; Writing—review & editing, A.D.H. and H.Y.; Visualization, A.D.H.; Funding acquisition, A.D.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a National Cancer Institute (NCI) Cancer Center Support Grant (CCSG) to Roswell Park Comprehensive Cancer Center (grant no. P30CA016056), and the following three NCI grants to Dr. Hutson: NRG Oncology Statistical and Data Management Center grant (grant no. U10CA180822); Cancer Immunoprevention Network (CIP-Net) grant (grant no. U24CA305623); Acquired Resistance to Therapy Network (ARTN) grant (grant no. U24CA274159).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data utilized to illustrate this work is publicly available.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Barry, D.A. *Statistical Methods in the Pharmaceutical Industry*; Marcel Dekker, Inc.: New York, NY, USA, 1990.
2. Peduzzi, P.; Concato, J.; Feinstein, A.R.; Holford, T.R. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J. Clin. Epidemiol.* **1995**, *48*, 1503–1510. [\[CrossRef\]](#)
3. Lin, N.X.; Logan, S.; Henley, W.E. Bias and Sensitivity Analysis When Estimating Treatment Effects from the Cox Model with Omitted Covariates. *Biometrics* **2013**, *69*, 850–860. [\[CrossRef\]](#)
4. Shao, Y.; Ye, Z.; Zhang, Z. Exact test and exact confidence interval for the Cox model. *Stat. Med.* **2024**, *epub ahead of print*. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Campone, M.; De Laurentiis, M.; Jhaveri, K.; Hu, X.; Ladoire, S.; Patsouris, A.; Zamagni, C.; Cui, J.; Cazzaniga, M.; Cil, T.; et al. Vepdegestrant, a PROTAC Estrogen Receptor Degradar, in Advanced Breast Cancer. *N. Engl. J. Med.* **2025**, *epub ahead of print*. [\[CrossRef\]](#)
6. Turner, N.C.; Im, S.-A.; Saura, C.; Juric, D.; Loibl, S.; Kalinsky, K.; Schmid, P.; Loi, S.; Sunpaweravong, P.; Musolino, A.; et al. Inavolisib-Based Therapy in PIK3CA-Mutated Advanced Breast Cancer. *N. Engl. J. Med.* **2024**, *391*, 1584–1596. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Janjigian, Y.Y.; Al-Batran, S.E.; Wainberg, Z.A.; Muro, K.; Molena, D.; Van Cutsem, E.; Hyung, W.J.; Wyrwicz, L.; Oh, D.Y.; Omori, T.; et al. Perioperative Durvalumab in Gastric and Gastroesophageal Junction Cancer. *N. Engl. J. Med.* **2025**, *epub ahead of print*. [\[CrossRef\]](#)
8. Wu, Y.L.; Dziadziuszko, R.; Ahn, J.S.; Barlesi, F.; Nishio, M.; Lee, D.H.; Lee, J.S.; Zhong, W.; Horinouchi, H.; Mao, W.; et al. Alectinib in Resected ALK-Positive Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2024**, *390*, 1265–1276. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Hoepfner, J.; Brunner, T.; Schmoor, C.; Bronsert, P.; Kulemann, B.; Claus, R.; Utzolino, S.; Izbicki, J.R.; Gockel, I.; Gerdes, B.; et al. Perioperative Chemotherapy or Preoperative Chemoradiotherapy in Esophageal Cancer. *N. Engl. J. Med.* **2025**, *392*, 323–335. [\[CrossRef\]](#)
10. Zhou, C.; Solomon, B.; Loong, H.H.; Park, K.; Perol, M.; Arriola, E.; Novello, S.; Han, B.; Zhou, J.; Ardizzoni, A.; et al. First-Line Selpercatinib or Chemotherapy and Pembrolizumab in RET Fusion-Positive NSCLC. *N. Engl. J. Med.* **2024**, *389*, 1839–1850. [\[CrossRef\]](#)
11. Herrera, A.F.; LeBlanc, M.; Castellino, S.M.; Li, H.; Rutherford, S.C.; Evens, A.M.; Davison, K.; Punnett, A.; Parsons, S.K.; Ahmed, S.; et al. Nivolumab plus AVD in Advanced-Stage Classic Hodgkin’s Lymphoma. *N. Engl. J. Med.* **2024**, *391*, 1379–1389. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Cheng, Y.; Spigel, D.R.; Cho, B.C.; Laktionov, K.K.; Fang, J.; Chen, Y.; Zenke, Y.; Lee, K.H.; Wang, Q.; Navarro, A.; et al. Durvalumab after Chemoradiotherapy in Limited-Stage Small-Cell Lung Cancer. *N. Engl. J. Med.* **2024**, *391*, 1313–1327. [\[CrossRef\]](#)
13. Munir, T.; Cairns, D.A.; Bloor, A.; Allsup, D.; Cwynarski, K.; Pettitt, A.; Paneesha, S.; Fox, C.P.; Eyre, T.A.; Forconi, F.; et al. Chronic Lymphocytic Leukemia Therapy Guided by Measurable Residual Disease. *N. Engl. J. Med.* **2024**, *390*, 326–337. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Zhou, C.; Tang, K.J.; Cho, B.C.; Liu, B.; Paz-Ares, L.; Cheng, S.; Kitazono, S.; Thiagarajan, M.; Goldman, J.W.; Sabari, J.K.; et al. Amivantamab plus Chemotherapy in NSCLC with EGFR Exon 20 Insertions. *N. Engl. J. Med.* **2024**, *389*, 2039–2051. [\[CrossRef\]](#) [\[PubMed\]](#)

15. Eskander, R.N.; Sill, M.W.; Beffa, L.; Moore, R.G.; Hope, J.M.; Musa, F.B.; Mannel, R.; Shahin, M.S.; Cantuaria, G.H.; Girda, E.; et al. Pembrolizumab plus Chemotherapy in Advanced Endometrial Cancer. *N. Engl. J. Med.* **2023**, *388*, 2159–2170. [[CrossRef](#)] [[PubMed](#)]
16. Powles, T.; Valderrama, B.P.; Gupta, S.; Bedke, J.; Kikuchi, E.; Hoffman-Censits, J.; Iyer, G.; Vulsteke, C.; Park, S.H.; Shin, S.J.; et al. Enfortumab Vedotin and Pembrolizumab in Untreated Advanced Urothelial Cancer. *N. Engl. J. Med.* **2024**, *390*, 875–888. [[CrossRef](#)]
17. Fakih, M.G.; Salvatore, L.; Esaki, T.; Modest, D.P.; Lopez-Bravo, D.P.; Taieb, J.; Karamouzis, M.V.; Ruiz-Garcia, E.; Kim, T.W.; Kuboki, Y.; et al. Sotorasib plus Panitumumab in Refractory Colorectal Cancer with Mutated KRAS G12C. *N. Engl. J. Med.* **2023**, *389*, 2125–2139. [[CrossRef](#)]
18. Powles, T.; Catto, J.W.F.; Galsky, M.D.; Al-Ahmadie, H.; Meeks, J.J.; Nishiyama, H.; Vu, T.Q.; Antonuzzo, L.; Wiechno, P.; Atduiev, V.; et al. Perioperative Durvalumab with Neoadjuvant Chemotherapy in Operable Bladder Cancer. *N. Engl. J. Med.* **2024**, *391*, 1773–1786. [[CrossRef](#)]
19. Jhaveri, K.L.; Im, S.A.; Saura, C.; Loibl, S.; Kalinsky, K.; Schmid, P.; Loi, S.; Thanopoulou, E.; Shankar, N.; Jin, Y.; et al. Overall Survival with Inavolisib in PIK3CA-Mutated Advanced Breast Cancer. *N. Engl. J. Med.* **2025**, *epub ahead of print*. [[CrossRef](#)]
20. Fizazi, K.; Piulats, J.M.; Reaume, M.N.; Ostler, P.; McDermott, R.; Gingerich, J.R.; Pintus, E.; Sridhar, S.S.; Bambury, R.M.; Emmenegger, U.; et al. Rucaparib or Physician’s Choice in Metastatic Prostate Cancer. *N. Engl. J. Med.* **2023**, *388*, 719–732. [[CrossRef](#)]
21. Wakelee, H.; Liberman, M.; Kato, T.; Tsuboi, M.; Lee, S.-H.; Gao, S.; Chen, K.-N.; Dooms, C.; Majem, M.; Eigendorff, E.; et al. Perioperative Pembrolizumab for Early-Stage Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2023**, *389*, 491–503. [[CrossRef](#)]
22. Mirza, M.R.; Chase, D.M.; Slomovitz, B.M.; Christensen, R.D.; Novak, Z.; Black, D.; Gilbert, L.; Sharma, S.; Valabrega, G.; Landrum, L.M.; et al. Dostarlimab for Primary Advanced or Recurrent Endometrial Cancer. *N. Engl. J. Med.* **2023**, *388*, 2145–2158. [[CrossRef](#)]
23. Cho, B.C.; Lu, S.; Felip, E.; Spira, A.I.; Girard, N.; Lee, J.S.; Lee, S.H.; Ostapenko, Y.; Danchaivijitr, P.; Liu, B. Amivantamab plus Lazertinib in Previously Untreated EGFR-Mutated Advanced NSCLC. *N. Engl. J. Med.* **2024**, *391*, 1486–1498. [[CrossRef](#)] [[PubMed](#)]
24. Slamon, D.; Lipatov, O.; Nowecki, Z.; McAndrew, N.; Kukiela-Budny, B.; Stroyakovskiy, D.; Yardley, D.A.; Huang, C.S.; Fasching, P.A.; Crown, J. Ribociclib plus Endocrine Therapy in Early Breast Cancer. *N. Engl. J. Med.* **2024**, *390*, 1080–1091. [[CrossRef](#)]
25. Hadoux, J.; Elisei, R.; Brose, M.S.; Hoff, A.O.; Robinson, B.G.; Gao, M.; Jarzab, B.; Isaev, P.; Kopeckova, K.; Wadsley, J.; et al. Phase 3 Trial of Selpercatinib in Advanced RET-Mutant Medullary Thyroid Cancer. *N. Engl. J. Med.* **2023**, *389*, 1851–1861. [[CrossRef](#)]
26. Schrag, D.; Shi, Q.; Weiser, M.R.; Gollub, M.J.; Saltz, L.B.; Musher, B.L.; Goldberg, J.; Al Baghdadi, T.; Goodman, K.A.; McWilliams, R.R.; et al. Preoperative Treatment of Locally Advanced Rectal Cancer. *N. Engl. J. Med.* **2023**, *389*, 322–334. [[CrossRef](#)]
27. Turner, N.C.; Oliveira, M.; Howell, S.J.; Dalenc, F.; Cortes, J.; Moreno, H.L.G.; Hu, X.; Jhaveri, K.; Krivorotko, P.; Loibl, S.; et al. Capivasertib in Hormone Receptor-Positive Advanced Breast Cancer. *N. Engl. J. Med.* **2023**, *388*, 2058–2070. [[CrossRef](#)] [[PubMed](#)]
28. Plante, M.; Kwon, J.S.; Ferguson, S.; Samouëlian, V.; Ferron, G.; Maulard, A.; de Kroon, C.; Van Driel, W.; Tidy, J.; Williamson, K. Simple versus Radical Hysterectomy in Women with Low-Risk Cervical Cancer. *N. Engl. J. Med.* **2024**, *390*, 819–829. [[CrossRef](#)] [[PubMed](#)]
29. Moore, K.N.; Angelergues, A.; Konecny, G.E.; Garcia, Y.; Banerjee, S.; Lorusso, D.; Lee, J.-Y.; Moroney, J.W.; Colombo, N.; Roszak, A.; et al. Mirvetuximab Soravtansine in FRalpha-Positive, Platinum-Resistant Ovarian Cancer. *N. Engl. J. Med.* **2023**, *389*, 2162–2174. [[CrossRef](#)]
30. Courneya, K.S.; Vardy, J.L.; O’Callaghan, C.J.; Gill, S.; Friedenreich, C.M.; Wong, R.K.; Dhillon, H.M.; Coyle, V.; Chua, N.S.; Jonker, D.J.; et al. Structured Exercise after Adjuvant Chemotherapy for Colon Cancer. *N. Engl. J. Med.* **2025**, *epub ahead of print*. [[CrossRef](#)]
31. Cascone, T.; Awad, M.M.; Spicer, J.D.; He, J.; Lu, S.; Sepesi, B.; Tanaka, F.; Taube, J.M.; Cornelissen, R.; Havel, L. Perioperative Nivolumab in Resectable Lung Cancer. *N. Engl. J. Med.* **2024**, *390*, 1756–1769. [[CrossRef](#)]
32. Shitara, K.; Van Cutsem, E.; Gümüş, M.; Lonardi, S.; de la Fouchardière, C.; Coutzac, C.; Dekervel, J.; Hochhauser, D.; Shen, L.; Mansoor, W. Trastuzumab Deruxtecan or Ramucirumab plus Paclitaxel in Gastric Cancer. *N. Engl. J. Med.* **2025**, *epub ahead of print*. [[CrossRef](#)]
33. Prager, G.W.; Taieb, J.; Fakih, M.; Ciardiello, F.; Van Cutsem, E.; Elez, E.; Cruz, F.M.; Wyrwicz, L.; Stroyakovskiy, D.; Pápai, Z.; et al. Trifluridine-Tipiracil and Bevacizumab in Refractory Metastatic Colorectal Cancer. *N. Engl. J. Med.* **2023**, *388*, 1657–1667. [[CrossRef](#)] [[PubMed](#)]
34. Kunkler, I.H.; Williams, L.J.; Jack, W.J.; Cameron, D.A.; Dixon, J.M. Breast-Conserving Surgery with or without Irradiation in Early Breast Cancer. *N. Engl. J. Med.* **2023**, *388*, 585–594. [[CrossRef](#)]
35. Hutson, A.; Yu, H. Strategies to Boost Statistical Efficiency in Randomized Oncology Trials with Primary Time-to-Event Endpoints. *Stat. Methods Med. Res.* **2025**, *34*, 1534–1552 [[CrossRef](#)] [[PubMed](#)]
36. Kempthorne, O. The Randomization Theory of Experimental Inference. *J. Am. Stat. Assoc.* **1955**, *50*, 946–967 [[PubMed](#)]
37. Cox, D.R.; Oakes, D. *Analysis of Survival Data*; Chapman and Hall/CRC: New York, NY, USA, 1984.

38. Therneau, M.T.; Grambsch, P.M.; Fleming, T.R. Martingale-Based Residuals for Survival Models. *Biometrika* **1990**, *77*, 147–160. [[CrossRef](#)]
39. Rosenbaum, P.R. Covariance Adjustment in Randomized Experiments and Observational Studies. *Stat. Sci.* **2002**, *17*, 286–327. [[CrossRef](#)]
40. Sweeney, C.J.; Chen, Y.-H.; Carducci, M.; Liu, G.; Jarrard, D.F.; Eisenberger, M.; Wong, Y.-N.; Hahn, N.; Kohli, M.; Cooney, M.M.; et al. Chemohormonal therapy in metastatic hormone-sensitive prostate cancer. *N. Engl. J. Med.* **2015**, *373*, 737–746. [[CrossRef](#)]
41. Salgia, R.; Stille, J.R.; Weaver, R.W.; McCleod, M.; Hamid, O.; Polzer, J.; Roberson, S.; Flynt, A.; Spigel, D.R. A randomized phase II study of LY2510924 and carboplatin/etoposide versus carboplatin/etoposide in extensive-disease small cell lung cancer. *Lung Cancer* **2017**, *105*, 7–13. [[CrossRef](#)]
42. Ryu, S.Y.; Deng, W.; Albuquerque, K.; Koh, W.J.; Mayadev, J.; Heugel, A.; Kim, B.J.; Kim, D.Y.; Cho, C.H.; Kim, J.W.; et al. Randomized Phase III Trial of Adjuvant Radiation Versus Chemoradiation in Intermediate-Risk, Early-Stage Cervical Cancer Following Radical Hysterectomy and Lymphadenectomy: Results from NRG Oncology/GOG-263/KGOG 1008. *Ann. Oncol.* **2025**, *in press*. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.