Online Learning with Survival Data

Arielle Anderer
Cornell Johnson Graduate School of Management

Joint work with Hamsa Bastani and John Silberholz

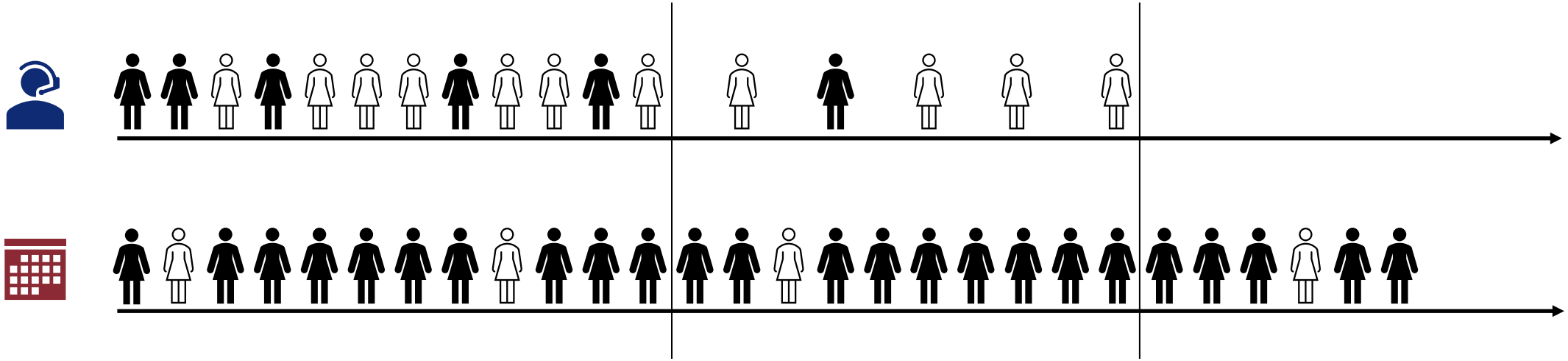
Johnson
Cornell
SC Johnson College of Business

Wharton
University of Pennsylvania

Robert H. Smith
School of Business

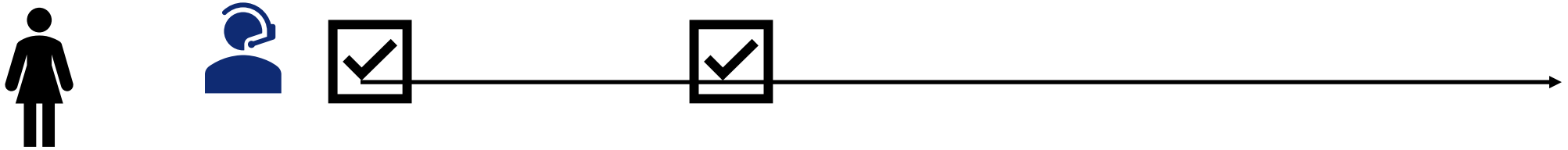
Adaptive Recruitment

- Typically assign patients to intervention with equal probability
- Increasing interest in rolling out best arm after experimentation
  - Test & roll in marketing (Feit & Berman 2019)
  - Quality improvement initiatives in healthcare and elsewhere
- Interest in adapting assignment probabilities over time
- Ex: Cancer Screening Outreach

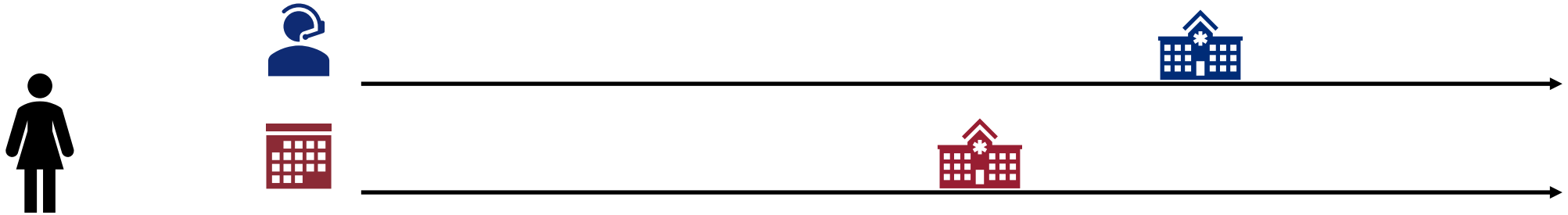


Traditional Multi-armed bandit

- You have different treatments available BUT you don't know which is best
- Each decision point t:
  - Patient t shows up & you assign them to treatment i
  - Person will experience an "outcome"
    - Outcome usually observed IMMEDIATELY after assignment
    - Sometimes observed after delay which is INDEPENDENT of arm reward

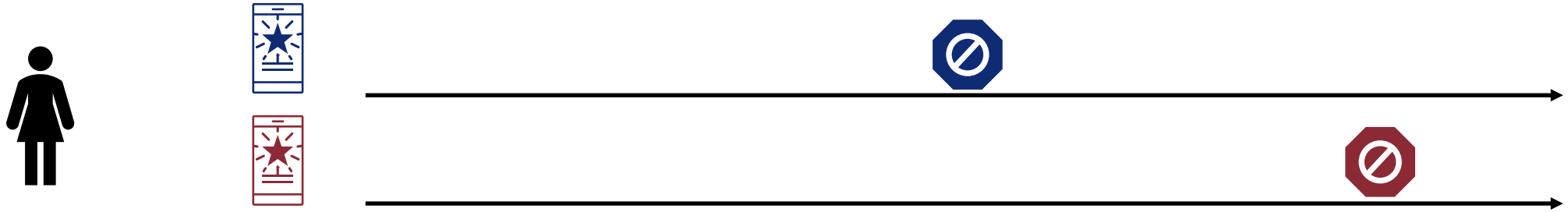


Time-to-Event Outcomes



- Healthcare:

  ex) Time to screening, progression free survival (> 60% of oncology outcomes![1])



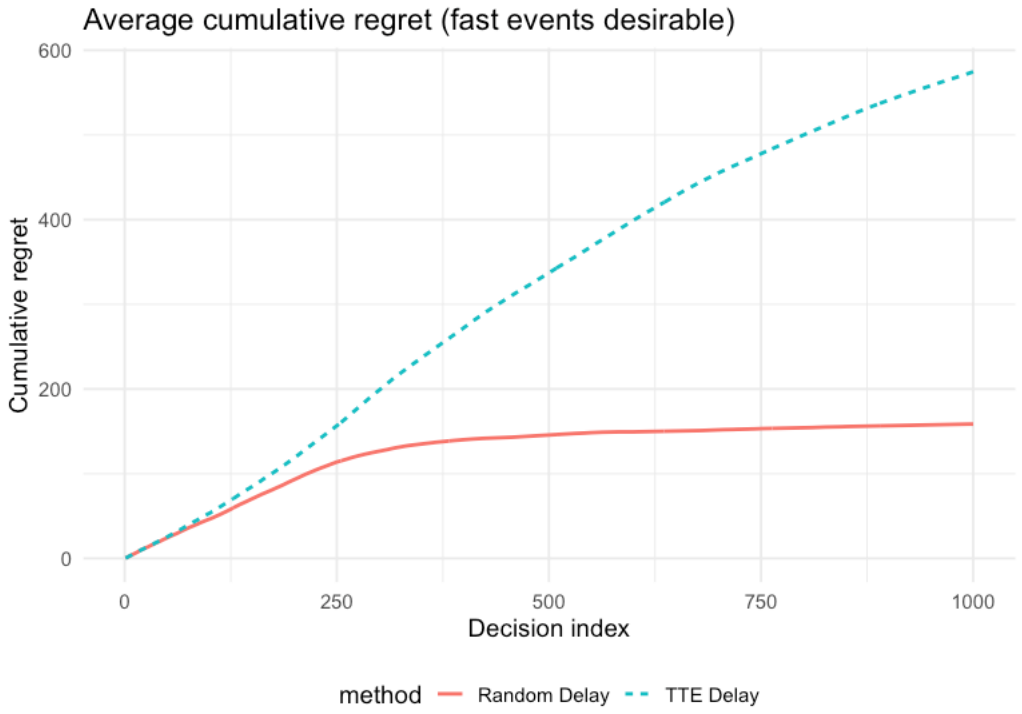
- Marketing, Human Resources, Manufacturing:

  ex) Churn, subscription cancellation, employee turnover, machine failure

1. Perlmutter et al. (2017)

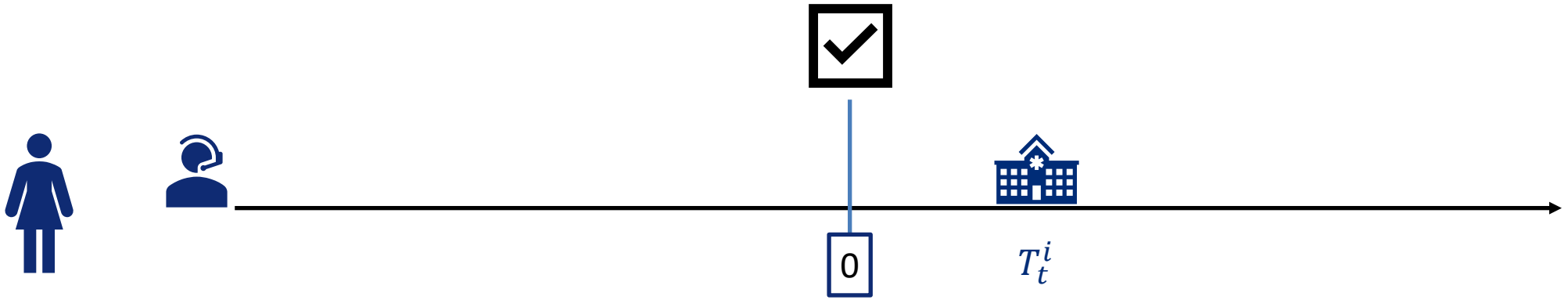
Naïve, Typical Methods for Delay

- Estimate average time-to-event
  - Ignore those without event
- Recent allocations bias estimate low
  - Only possibilities: very short event time or no event seen yet
- Can lead to pre-commitment to the wrong arm when minimizing time
  - Initial allocations introduce bias that encourages further allocations



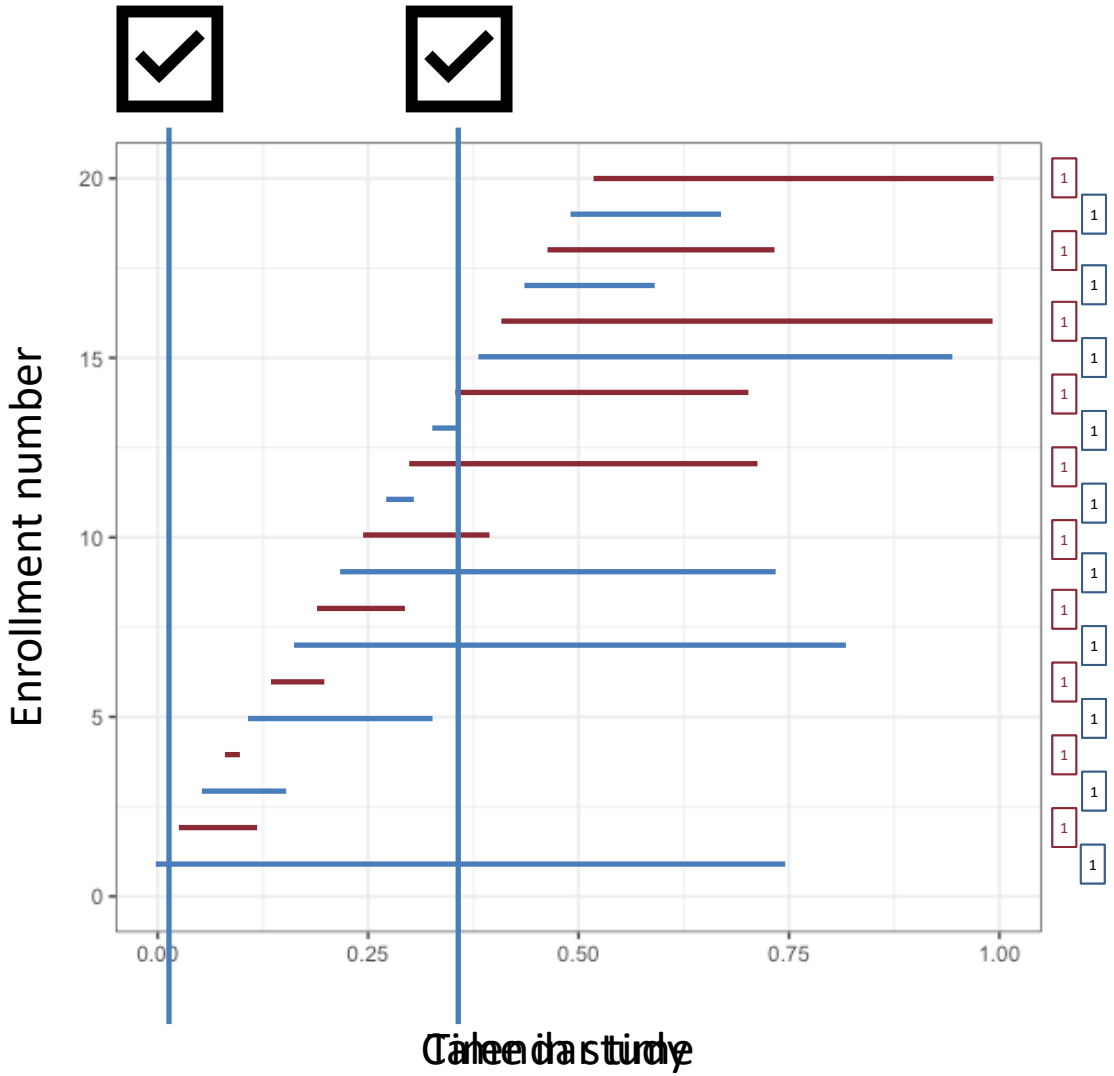
Current Methods for Dealing with Survival Data

- Dichotomizing:
  - Pick a time (e.g. 6 months)
  - Measurement = did patient get tested by this time
    - 1: was tested within 6 months
    - 0: was NOT tested within 6 months



Issues with Dichotomization

- Granularity of information
- How do we choose cutpoint?
  - Not long enough – uninformative!
  - Wait too long – uninformative!
  - Check too many thresholds – multiple hypothesis testing!



Contributions of Paper

- **Algorithm Design.**

  How do we leverage ideas from survival analysis literature to design an effective bandit algorithm that works well with time-to-event data?

- **Theoretical Guarantees.**

  Understanding when incorporating survival-based analysis is especially useful

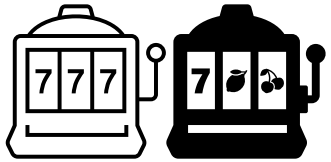
- **Evaluation.**

  Simulation of our algorithm numerically on cervical cancer screening intervention data

Literature Review

Survival Analysis

Cox (1972)
Andersen & Gil (1982)
Sellke & Siegmund (1983)
Machin et al (2006)
Luo et al. (2013)

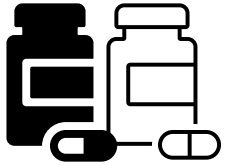


Multi-Armed Bandits

Thompson (1933)
Lai & Robbins (1985)
Rusmevichientong & Tsitsiklis (2010)

Joulani et al. (2013)
Liu et al. (2019)
Keyvanshokooh et al. (2019)
Wager & Wu (2022)

Wager & Xu (2021)
Fan & Glynn (2021)
Hirano & Porter (2021)
Kalvit & Zeevi (2021)



(Adaptive) Experimentation

Cheung et al. (2006)
Berry et al. (2010)
Ahuja and Birge (2016)
Bertsimas et al. (2016)
Kouvelis et al. (2017)
Chick et al. (2017)
Chick et al. (2018)
Besbes et al. (2019)
Corcoran et al. (2019)
Bastani & Bayati (2020)
Anderer et al. (2022)

Contributions of Paper

- **Bandit Algorithm Design.**

  How do we leverage ideas from survival analysis literature to design an effective bandit algorithm that works well with time-to-event data?

- **Theoretical Guarantees.**

  Understanding when incorporating survival-based analysis is especially useful

- **Evaluation.**

  Simulation of our algorithm numerically on cervical cancer screening intervention data

How to measure efficacy: Hazard ratio

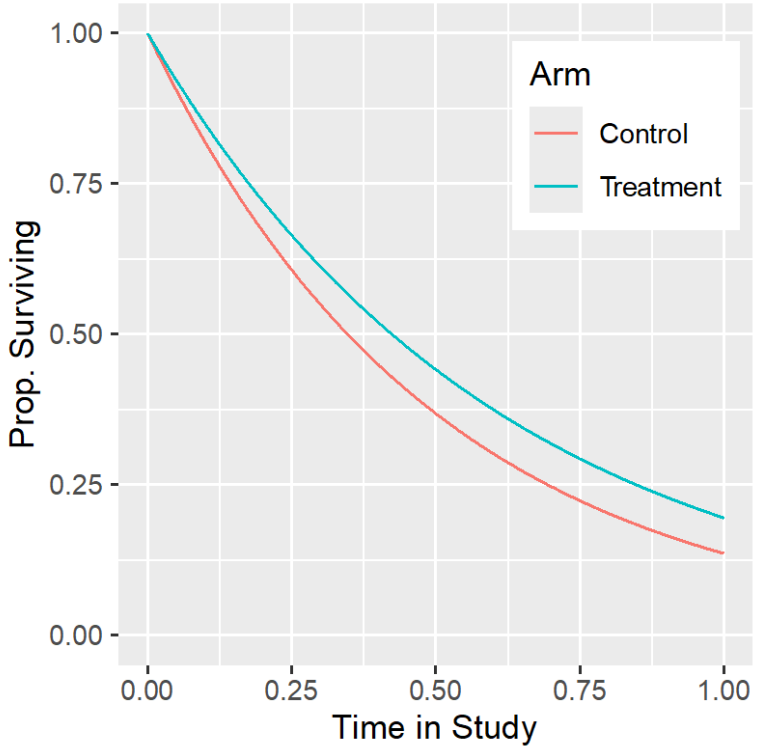
- Hazard rate $h_i(t)$: event rate among $N_i(t)$ at-risk individuals in arm i

$$h_i(t) = \lim_{\delta \to 0} \frac{\#\ events\ in\ [t, t+\delta]}{\delta \cdot N_i\ (t)}$$

- Hazard ratio: Analyze the efficacy of treatment as compared to the control

$$\frac{h_1(t)}{h_0(t)} = e^{\beta}$$

  - Proportional hazards (PH) assumption: constant ratio $HR(t) = e^{\beta}$

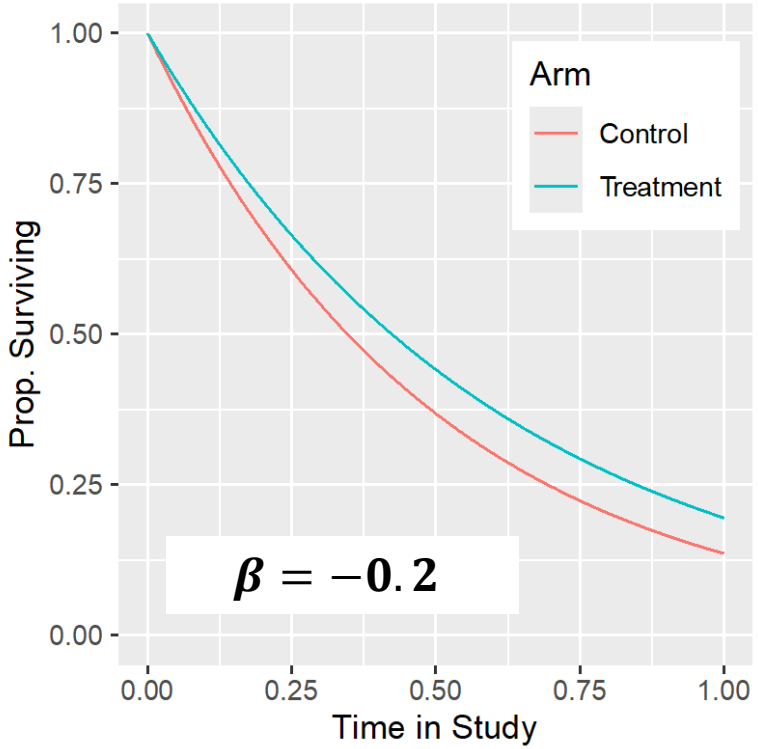


Survival Analysis: Cox proportional hazard model

- Hazard ratio: Analyze the efficacy of treatment as compared to the control

$$\frac{h_1(t)}{h_0(t)} = e^{\beta}$$

Under proportional hazards



- Estimate coefficient β (log hazard ratio)

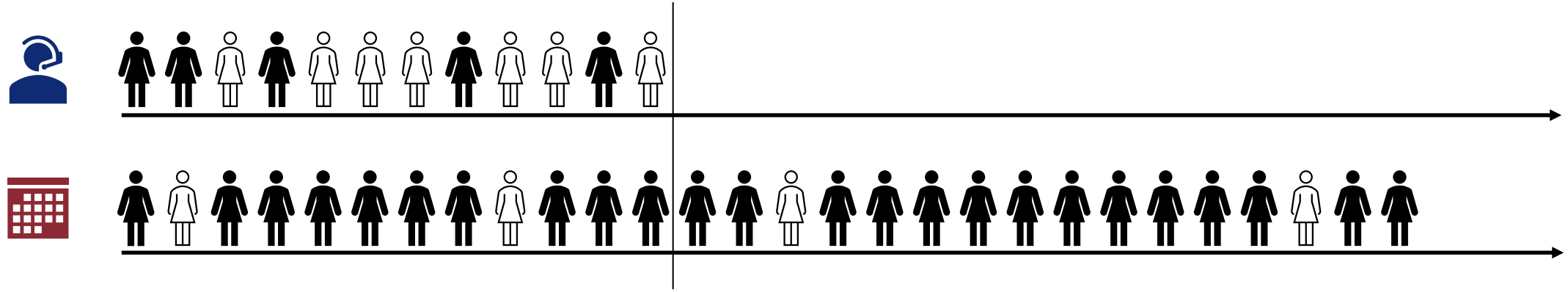
$\beta < 0$

$\beta \geq 0$

– Semi-parametric Cox PH model estimates β

Survival Explore-Then-Commit (ETC)

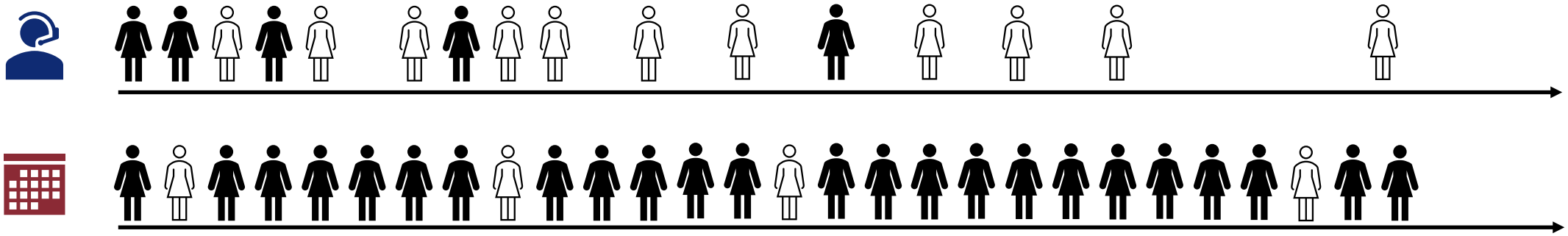
- While estimator $(\hat{\beta}, \hat{\sigma})$ has not reached a target precision v_n^* :
  - Perform 50/50 random allocation between arms
- Once estimator $(\hat{\beta}, \hat{\sigma})$ reaches target precision v_n^*:
  - If $\hat{\beta} < 0$, allocate to control arm forever*
  - If $\hat{\beta} > 0$, allocate to treatment arm forever



*If quicker time-to-event outcomes preferred. In other scenarios, swap allocation rule

Survival Thompson Sampling (TS)

- Each arrival, draw outcome $\tilde{\beta} \sim N\left(\hat{\beta}, (\hat{\sigma})^2\right)$ from current estimator
  - If $\tilde{\beta} < 0$, allocate new arrival to the control arm*
  - If $\tilde{\beta} > 0$, allocate new arrival to the treatment arm



*If quicker time-to-event outcomes preferred. In other scenarios, swap allocation rule

Contributions of Paper

- **Bandit Algorithm Design.**

  How do we leverage ideas from survival analysis literature to design an effective bandit algorithm that works well with time-to-event data?

- **Theoretical Guarantees.**

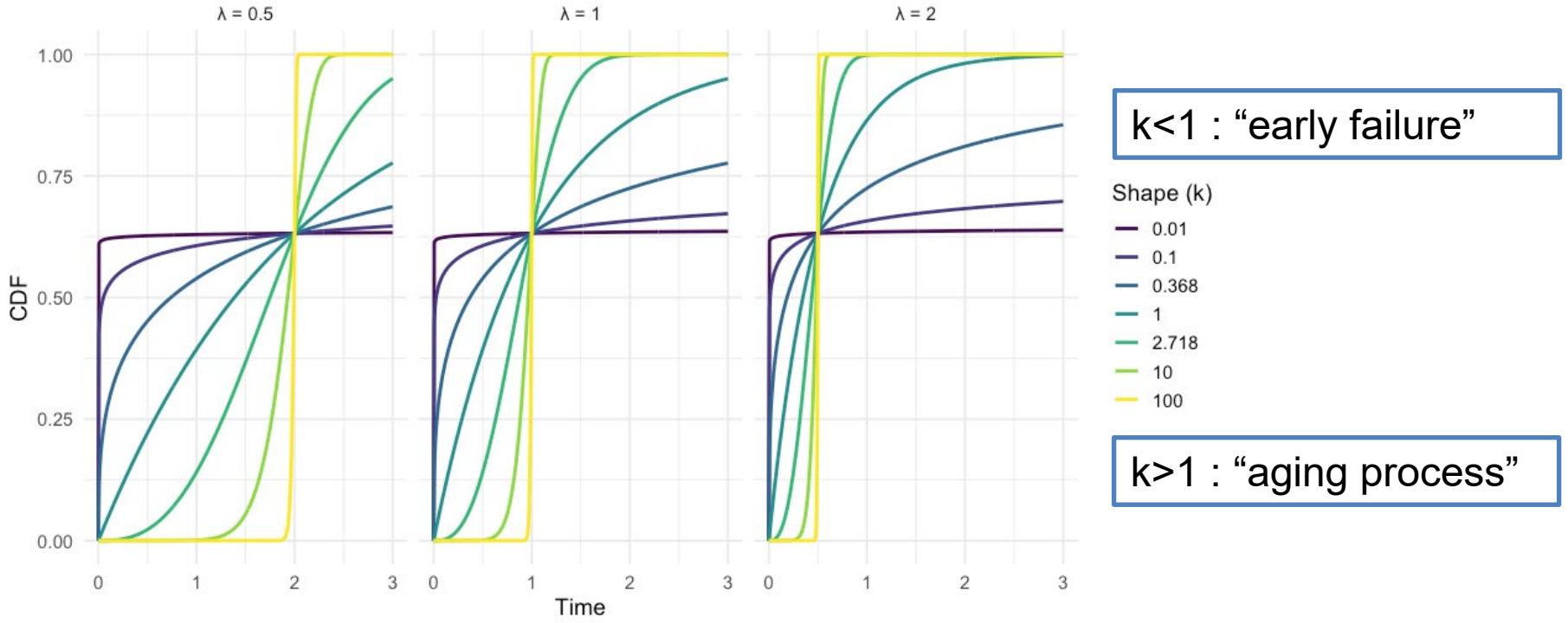
  Understanding when incorporating survival-based analysis is especially useful

- **Evaluation.**

  Simulation of our algorithm numerically on cervical cancer screening intervention data

Theoretical Analysis of Expected Regret

- Theoretical analysis setting
  - **Algorithm**: Explore-then-commit (ETC) for dichotomized and survival analysis
  - **Enrollment**: Uniform over unit-length time horizon
  - **Survival**: Weibull distributed event times with shape k and rate λ (control) and $\lambda e^{\beta/k}$ (treatment)



k<1 : “early failure”

k>1 : “aging process”

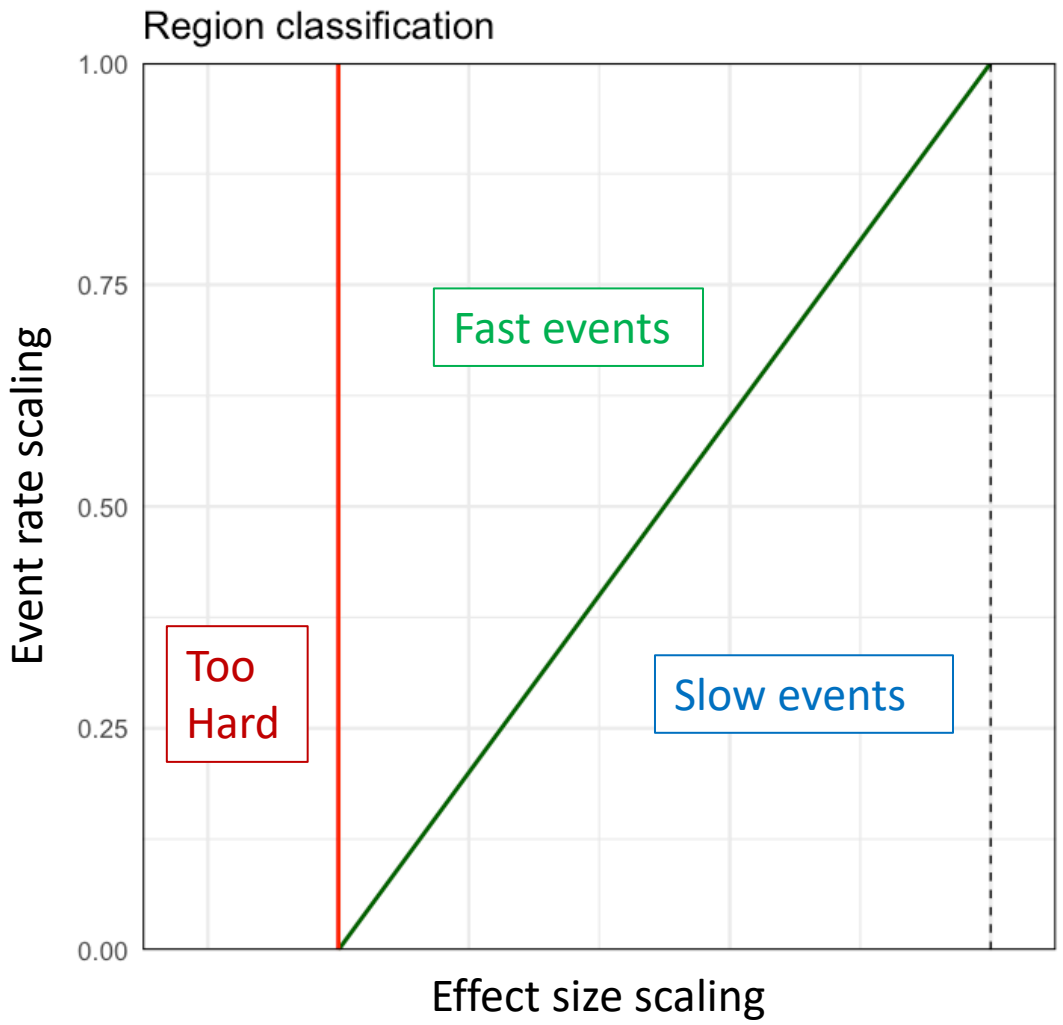
Theoretical Analysis of Expected Regret

- Theoretical analysis setting
  - **Algorithm**: Explore-then-commit (ETC) for dichotomized and survival analysis
  - **Enrollment**: Uniform over unit-length time horizon
  - **Survival**: Weibull distributed event times with shape k and rate λ (control) and $\lambda e^{\beta/k}$ (treatment)
  - **Performance metric**: expected regret R_n^{dich}, R_n^{surv} (# allocations to weaker arm)

- Theoretical results are for large-n regime (many patients in study)
  - Problem difficulty allowed to vary as instance size n grows:
    - Diminishing effect size: $\beta_n = \beta_0 n^s$ for $s \leq 0$
    - Event speed relative to horizon: $\lambda_n = \lambda_0 n^\alpha$ for $\alpha \geq 0$

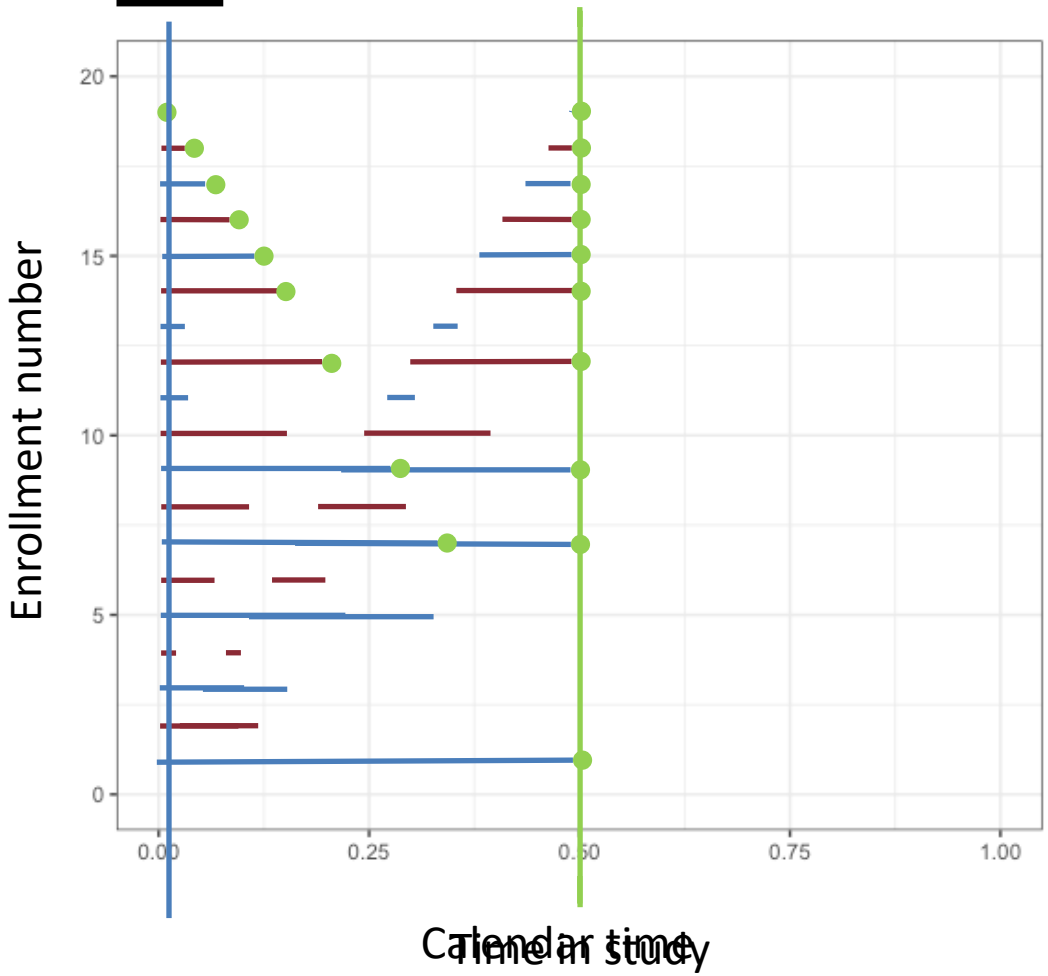
Regions

- Event rate scaling captures different regimes
- Slow Events Regime
  - Event arrival rate small vs. exploration time
  - Decisions under significant censoring
  - Cancer screening
  - Long-term customer churn
- Fast events regime
  - Event arrival rate large vs. exploration time
  - Little censoring
  - Short-term website dwell time



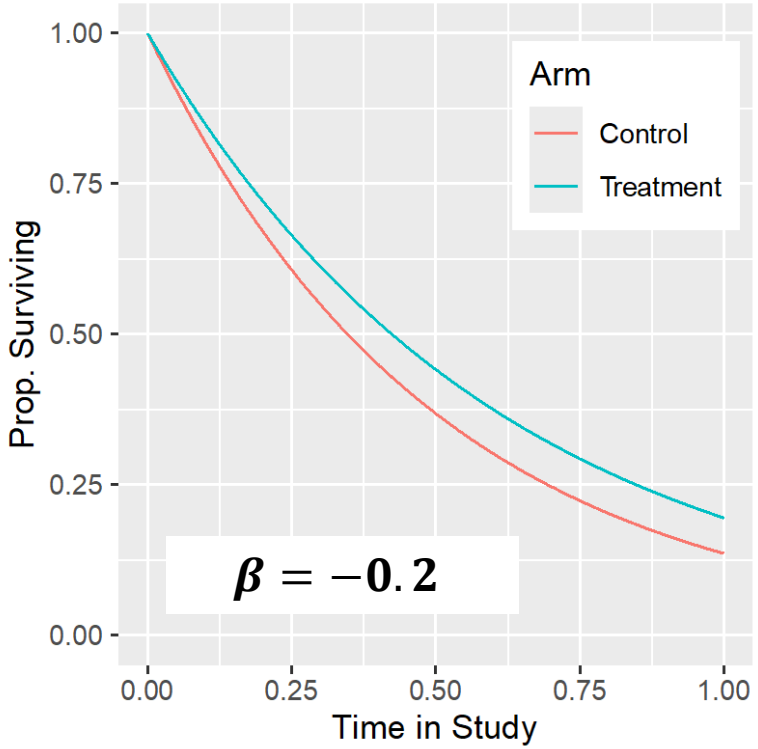
Choice of Cutpoint τ_n (Dichotomized) ☑

- Slow Events Regime
  - Cutpoint τ_n balances data quality vs data latency
    - Too small: miss late events
    - Too large: miss events from recent enrollees
  - Optimal choice: $\frac{\tau_n}{m_n} = \frac{k}{k+1}$
    - Early failure process: informative events arrive quickly -> small τ_n
    - Fatigue process: information accumulates slowly -> large τ_n



Choice of Cutpoint τ_n (Dichtomized)

- Fast events regime
  - No longer worried about censoring
  - τ_n often set at MEDIAN (50th percentile)
    - Common practice
    - Attempt to maximize variance
  - Optimal choice τ_n= 79.7th percentile!
    - Treatment effects accumulate over time
    - Waiting longer → larger gap between arms!
    - Waiting too long → event probability 1
  - Payoff of optimizing cutpoint: 35% gain relative to median

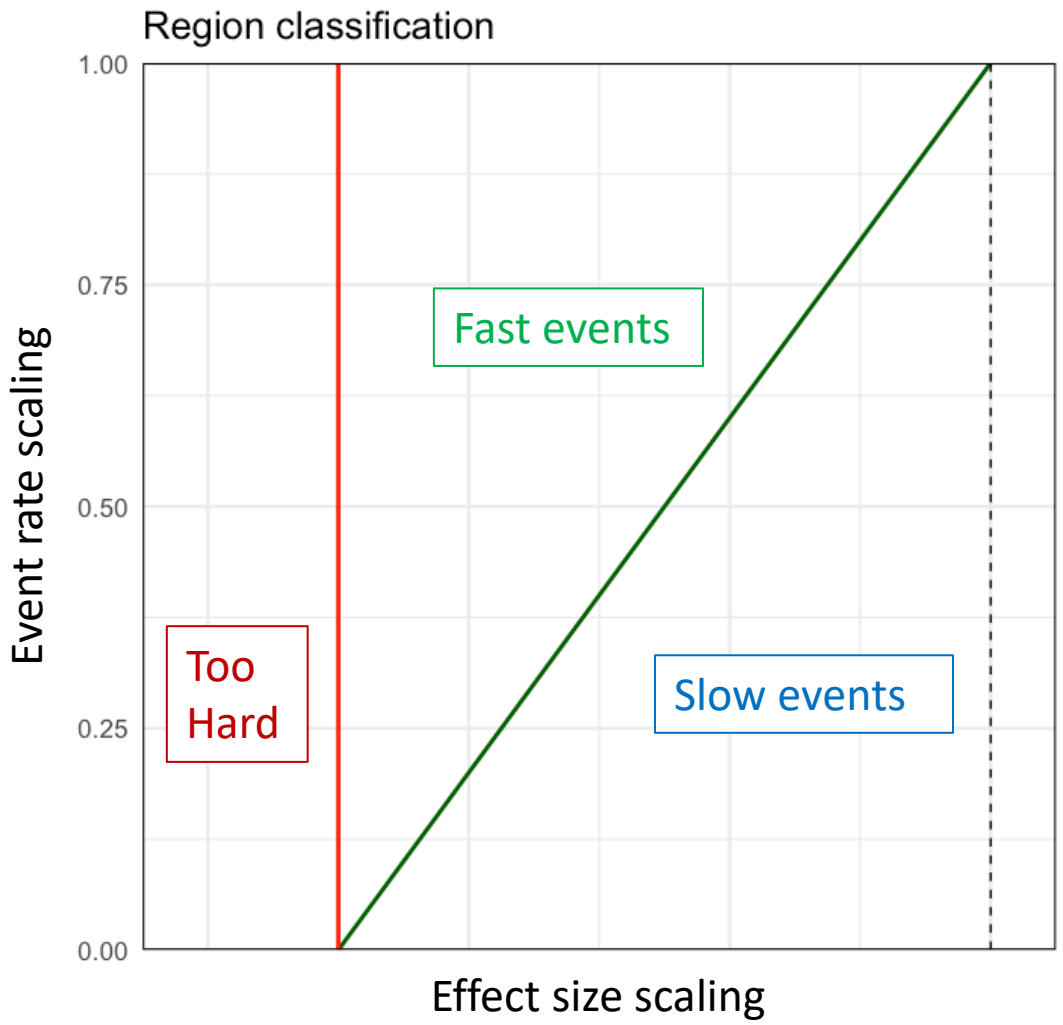


Comparison of Regret

- Thms 1 & 2 (informal):

$$R_n^{dich} \sim \begin{cases} \left(\frac{k+1}{k}\right)^{\frac{k}{k+1}} R_n^{surv} & -1-2s < \alpha < 1+2s \\ 1.544\, R_n^{surv} & \alpha \geq 1+2s\,, s > -1/2 \\ \mathrm{R}_\mathrm{n}^{surv} & o.w.\,(both\ linear) \end{cases}$$

  - The Survival Bandit algorithm shows sizable large-n improvements in regret (many cases > 41%) vs. the Dichotomized Bandit algorithm
  - Survival Analysis especially helpful with quick events



Regret as a function of Weibull Shape Parameter k

- Event rate scaling captures different regimes
- Slow Events Regime
  - Intermediate k: dichotomization inefficient
    - Peak improvement of 45% @ k= 0.58
    - Improvement of 41% @ k = 1 (exponential)
  - Extreme k: structural shortcuts reduce inefficiency
    - Early failure (k → 0)

      Events occur immediately or never
    - Aging process (k →∞)

      Events cluster together
- Fast events regime
  - Shape parameter k does not matter!

Comparison Under Uncertainty

- In practice, event rate rarely known a priori
  - Multiplicative uncertainty by a factor of r in cumulative hazard:

$$\Lambda_n(t) = (t\lambda_0 n^\alpha)^k \in \left[\left(t\underline{\lambda}_0 n^\alpha\right)^k, \mathrm{r}\left(t\underline{\lambda}_0 n^\alpha\right)^k\right]$$

  - Equivalently:

$$\lambda_0 \in \left[\underline{\lambda}_0, \underline{\lambda}_0 r^{1/k}\right]$$

- Choose algorithm parameters to minimize worst-case regret

$$\max_{\lambda_0} \min_{d} R_n(d, \lambda_0)$$

  - Survival ETC: $d =$ target precision ν_n
  - Dichotomized ETC: $d =$ target precision ν_n and cutpoint τ_n

- Key benchmark: undeniable regret!

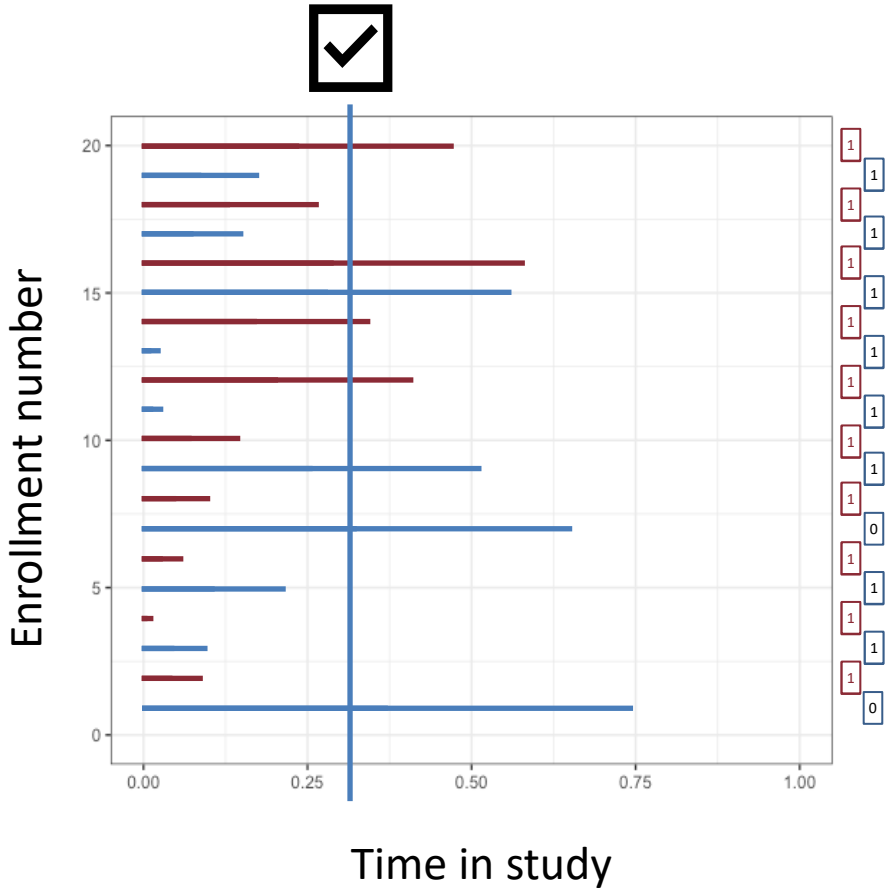
$$\max_{\lambda_0} \min_{d} R_n(d, \lambda_0) \leq \min_{d} \max_{\lambda_0} R_n(d, \lambda_0)$$

Uncertainty:

- **Survival Bandit**:
  - Uses exact event times
  - No need to choose optimal cutpoint
  - Optimal target precision ν_n^* does not rely on λ_0
  - **Takeaway:**

    Conservative choice of ν_n^* -> undeniable regret $\forall \lambda_0$
- **Dichotomized Bandit:**
  - Cutpoint τ_n must be fixed
  - Optimal τ_n^* depends on event rate!
  - Slow arrivals -> large threshold best
  - Fast arrivals -> small threshold best
  - **Takeaway:**

    Under uncertainty: may need to balance this!



Impact of Uncertainty:

Ratio of worst-case expected regret in “Fast events” regime

- Slow Regime:
  - Dichotomized sufficiently controls worst-case regret
  - Target lowest probable event rate -> achieve undeniable regret
- Fast Regime:
  - Optimal τ_n^* depends critically on event rate!
  - Want to dichotomize at specific percentile
  - Must select conservative τ_n^* -> substantial inefficiency

Contributions of Paper

- **Bandit Algorithm Design.**

  How do we leverage ideas from survival analysis literature to design an effective bandit algorithm that works well with time-to-event data?

- **Theoretical Guarantees.**

  Understanding when incorporating survival-based analysis is especially useful

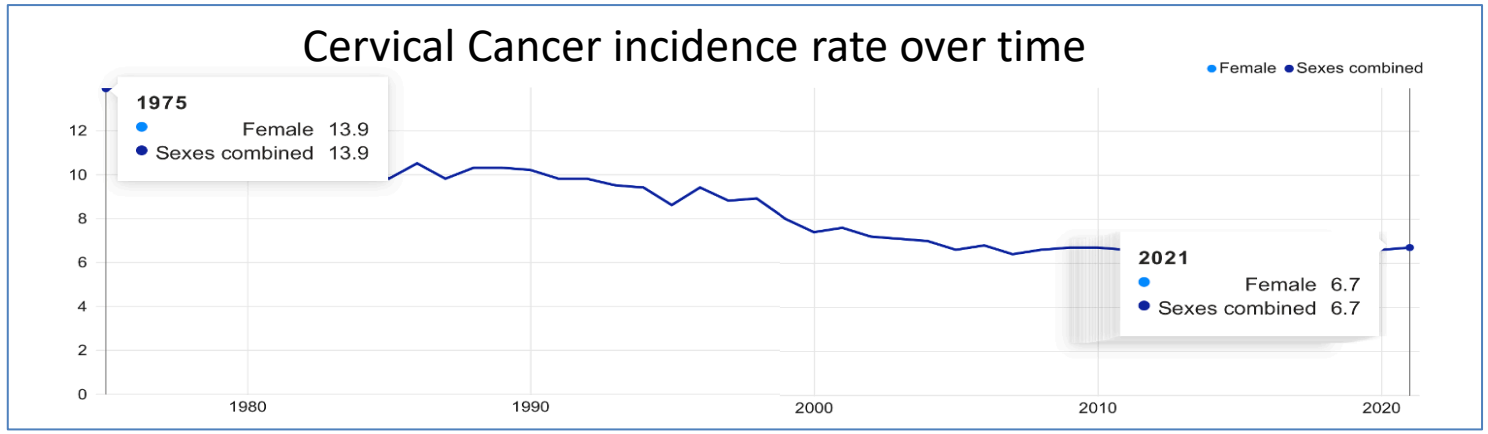
- **Evaluation.**

  Simulation of our algorithm numerically on cervical cancer screening intervention data

Cervical Cancer Screening motivation

In the United States for 2026:

- About 13,490 new cases of invasive cervical cancer will be diagnosed.
- About 4,200 women will die from cervical cancer.



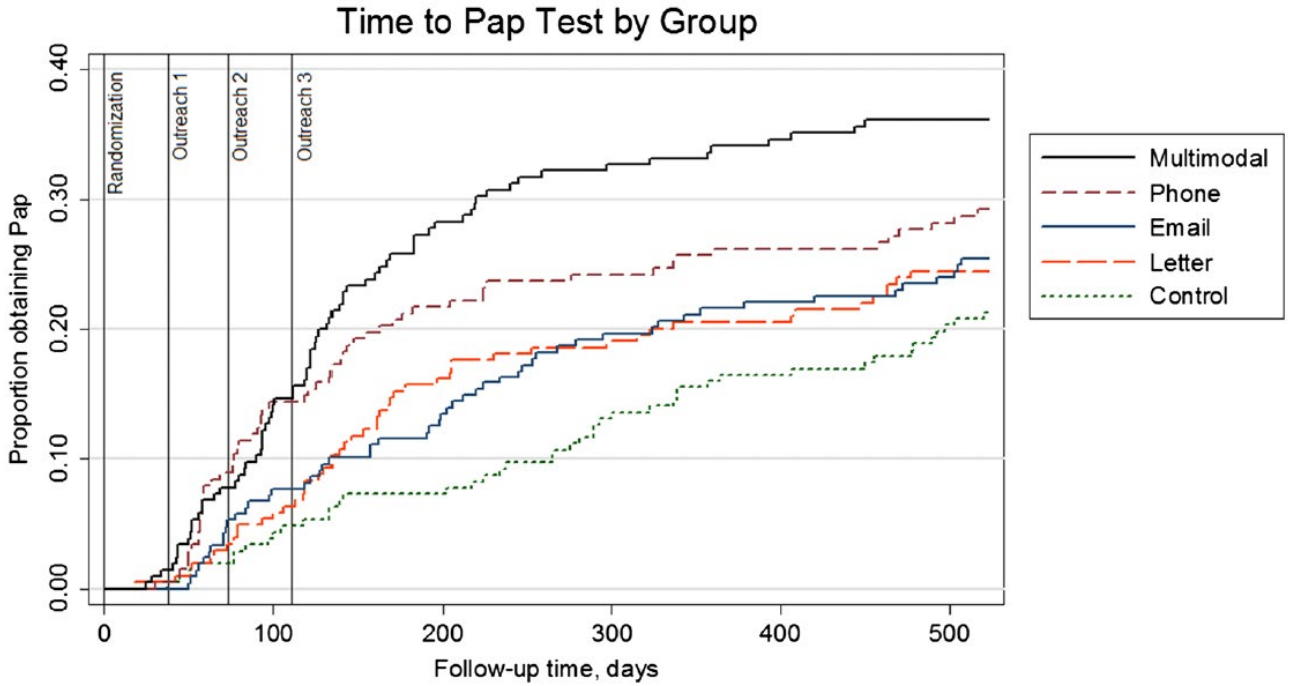
- 50-70% of cervical cancers diagnosed in women not screened in last 5 years
- When detected before spread, one of most successfully treated cancers

Increasing compliance with screening guidelines

- ~83% of women screened according to guidelines
- Public health interventions:
  - Outreach methods (mail, email, phone, multiple methods)
  - Outreach integration with scheduling/testing procedure
    - Access to appointment scheduling
    - Assigned “nurse navigator”
    - At-home HPV screening tests
- Need to design experiments to test efficacy of interventions!

Cervical Cancer Screening Simulation

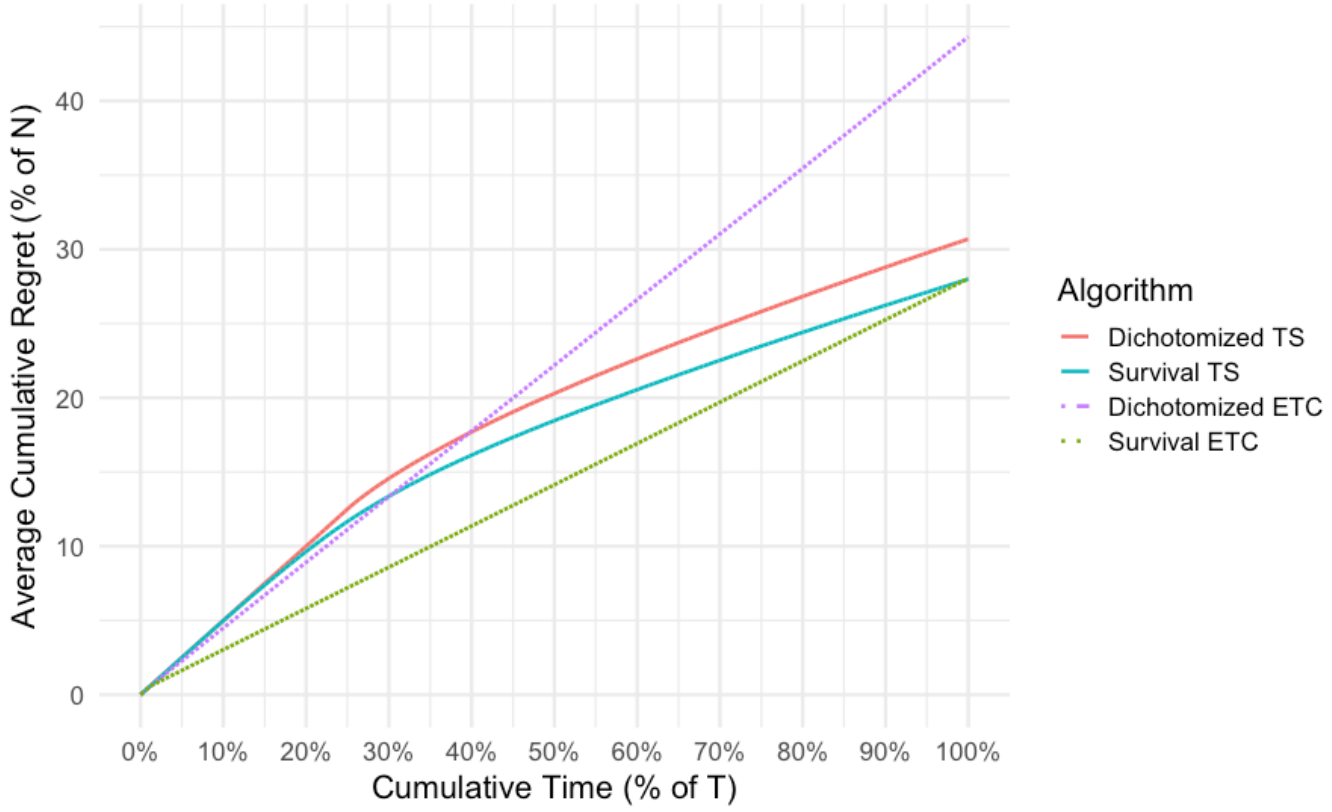
- Simulate using real-world data from behavioral interventions
- 18 pairs of control/treatment from 4 studies
  - Use Kaplan-Meier curves to extract participant event times
  - Simulate dichotomized Thompson Sampling vs. Survival Thompson Sampling
- Relaxes assumptions
  - Weibull event times
  - Proportional hazards
  - Large n



Regret over Time

	Survival ETC	Dichtomized ETC	Survival TS	Dichotomized TS
Terminal Regret	28.3	47.7	28	30.7

- ETC:
  - Survival 40.6% better than dichotomized
  - Limited exploration
- TS:
  - Survival 8.8% better than dichotomized
  - More exploration



Average (normalized) Regret at time T for various thresholds

Dichotomized TS regret

Conclusions

- Multiple advantages of “survival bandits” over dichotomization
  - Easier to parameterize
  - 41-54% improvements in regret in large-n regime under exponential event times
  - Over 30% improvements in regret for a wide variety of parameter settings
  - Can be even larger under event rate uncertainty
  - Strong simulated finite-sample performance

- Eager to identify more use cases and datasets to test out the proposed approach


THANK YOU
Any questions?
aanderer@cornell.edu
CORNELL UNIVERSITY
FOUNDED A.D. 1865
Johnson
Cornell
SC Johnson College of Business

Min-max Regret Under Uncertainty

Survival Bandit

$$\min_{\nu}\left\{\max_{(s,\alpha)\in U}\{R_n(\nu,s,\alpha)\}\right\}$$

$$\sim\begin{cases} n^{1/2-\alpha_0/2-s_0}\sqrt{\log(n)}\sqrt{\frac{2(1+\alpha_0+2s_0)}{\beta_0^2\lambda_0}}\Delta & for\ -1-2s_0<\alpha_0<1+2s_0 \\ n^{-2s_0}\log(n)\frac{4(1+2s_0)}{\beta_0^2\lambda_0} & for\ \alpha_0\geq 1+2s_0, s_0>1/2 \\ n\Delta/2 & o.w. \end{cases}$$

Binarized Bandit

$$\min_{\nu,\tau}\left\{\max_{(s,\alpha)\in U}\{R_n(\nu,\tau,s,\alpha)\}\right\}$$

$$\sim\begin{cases} n^{-2s_0}\sqrt{\log(n)}\sqrt{\frac{4(1+\alpha_0+2s_0)}{\beta_0^2\lambda_0}}\Delta & for\ -1-2s_0<\alpha_0<1+2s_0, \alpha_1<1+2s_0 \\ n^{-2s_0+(\alpha_1-\alpha_0)}\left[4\frac{(1+2s_0-(\alpha_1-\alpha_0))}{(\alpha_1-\alpha_0)\beta_0^2}\right]\Delta & for\ -1-2s_0<\alpha_0<1+2s_0, \frac{1+\alpha_0+2s_0}{2}\leq\alpha_1 \\ n^{-2s_0+(\alpha_1-\alpha_0)}\left[4\frac{(1+2s_0-(\alpha_1-\alpha_0))}{(\alpha_1-\alpha_0)\beta_0^2}\right]\Delta & for\ \alpha_0\geq 1+2s_0, s_0>1/2, \\ n\Delta/2 & o.w. \end{cases}$$

- Survival Bandit chooses the optimal ν for parameters (s_0,α_0), incurs same worst-case regret
- For cases where $(-1-2s_0<\alpha_0<1+2s_0\ \&\ \alpha_1\geq 1+2s_0)$ OR $\alpha_0\geq 1+2s_0$:

  Binarized Bandit must choose MORE conservative binarization time τ_n^*, which leads to higher worst-case regret by a factor of $n^{(\alpha_1-\alpha_0)}$
- If gap large enough $(\alpha_1>1+\alpha_0+2s_0)$: no sublinear regret