med fiz

Univerza v Ljubljani
Fakulteta za matematiko in fiziko

Evaluation of Lung SUV_{95} as a Candidate Surrogate Endpoint in Metastatic Melanoma

Zahra Alirezaei

Introduction

Clinical motivation

- Immune checkpoint inhibitors (ICI) improve survival in metastatic melanoma
- ICI-induced adverse events remain clinically important
- Imaging biomarkers may help identify toxicity risk

Main question: Can lung SUV_{95} substitute for the true clinical endpoint?

Introduction

What is a surrogate endpoint?

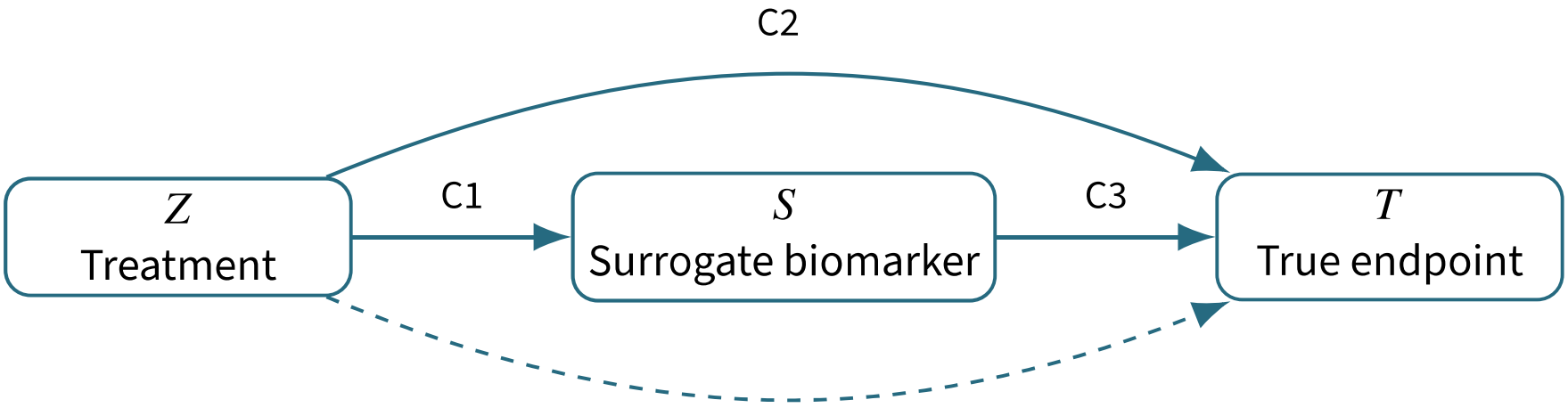
$$Z \rightarrow S \rightarrow T$$

- Z = treatment
- S = candidate surrogate biomarker
- T = true clinical endpoint

A valid surrogate captures the treatment pathway, so treatment adds no extra information once S is known.

Introduction

Surrogacy framework



Interpretation: a valid surrogate requires the treatment effect on T to be captured through S.

Introduction

Classical Prentice criteria

$$P(S|Z) \neq P(S) \qquad P(T|Z) \neq P(T)$$

$$P(T|S) \neq P(T) \qquad P(T|Z,S) = P(T|S)$$

- **C1:** treatment affects the surrogate
- **C2:** treatment affects the endpoint
- **C3:** the surrogate predicts the endpoint
- **C4:** treatment effect disappears after adjusting for the surrogate

All four conditions are required for classical surrogate validation.

Introduction

Example illustrating the four Prentice criteria

Criterion.	Data example
C1	Mean SUV_{95} by treatment:
Treatment A = 1.2	
Treatment B = 2.0	Treatment changes the surrogate
C2	Lung AE rate by treatment:
Treatment A = 10%	
Treatment B = 30%	Treatment changes the true endpoint
C3	Lung AE rate by SUV_{95} group:
Low SUV = 5%	
High SUV = 40%	The surrogate predicts the endpoint
C4	After conditioning on SUV_{95}:
Low SUV: A = 5%, B = 5%	
High SUV: A = 40%, B = 40%	Within each SUV group, treatment no longer changes AE risk

Key idea: C4 means that once the surrogate is known, the treatment effect on the endpoint disappears.

Introduction

Why Prentice alone is not enough

- Prentice is conceptually clear, but very strict
- A biomarker may predict outcome without mediating treatment effect
- Single-cohort studies need complementary analyses

Therefore, we combined Prentice with:

- Freedman proportion explained
- Wang–Taylor sensitivity analysis

Literature Review

Foundational surrogate endpoint frameworks

Paper	Context	Method	Key idea
Prentice (1989)	theoretical framework	4 criteria	Defines the pathway $Z \rightarrow S \rightarrow T$ required for surrogate validation
Freedman (1992)	clinical trials	proportion explained	Quantifies how much the surrogate attenuates the treatment effect
Wang & Taylor (2004)	general framework	risk-scale mediation	Evaluates mediation on the probability scale rather than coefficients alone
Ray (2009)	prostate cancer trials	Prentice implementation	Applies surrogate evaluation in a clinical survival setting
Heller (2015)	oncology trials	single-trial Prentice analysis	Evaluates candidate surrogates within randomized-trial data

Literature Review

Trial-level and meta-analytic validation frameworks

Paper	Context	Method	Key idea
Buyse (2000)	multiple cancer trials	meta-analytic validation	Correlates treatment effects on surrogate and true endpoint across trials
Burzykowski (2005)	oncology trials	surrogate threshold effect	Defines the minimum surrogate effect needed to predict true endpoint benefit
Daniels & Hughes	multiple trials	hierarchical Bayesian model	Jointly models treatment effects on surrogate and clinical endpoint
Alonso (2004)	clinical trials	canonical correlation	Evaluates association structure between surrogate and clinical outcomes

Literature Review

Position of the present study

- Surrogate-validation frameworks depend on data structure
- Trial-level validation requires multiple randomized trials
- Our dataset is a single cohort with few events

Therefore, we focused on:

- individual-level evaluation
- Prentice framework
- complementary analyses

Methods

Study design

- Metastatic melanoma cohort
- $n = 58$ patients
- 5 lung adverse events

$Z =$ treatment $\quad S =$ lung SUV_{95} $\quad T =$ lung adverse event

Primary analysis used **continuous SUV_{95}**.

Methods

Prentice framework and interpretation

Statistical models

C1: $S = \alpha_0 + \alpha_1 Z + \varepsilon$

C2: $\text{logit}(P(T = 1)) = \beta_0 + \beta_1 Z$

C3: $\text{logit}(P(T = 1)) = \gamma_0 + \gamma_1 S$

C4: $\text{logit}(P(T = 1)) = \delta_0 + \delta_1 Z + \delta_2 S$

$$PE = 1 - \frac{\beta_{adj}}{\beta_{unadj}} \qquad PE_{risk} = \frac{AA - AB}{AA - BB}$$

Interpretation

- α_1: treatment → biomarker
- β_1: treatment → outcome
- γ_1: biomarker → outcome
- δ_1: residual treatment effect
- δ_2: adjusted biomarker effect

Key idea

- β_{unadj} vs β_{adj}: effect reduction
- AA: full effect
- AB: treatment only
- BB: baseline

Results

Main Prentice results

Crit.	Model	Main result	Interpretation
C1	$S \sim Z$	coef $= -0.019$, $p = 0.923$	No evidence that treatment affects SUV_{95}
C2	$T \sim Z$	OR $= 1.96$, $p = 0.575$	No significant treatment effect on adverse events
C3	$T \sim S$	OR ≈ 295, $p = 0.004$	SUV_{95} strongly predicts adverse-event risk
C4	$T \sim Z + S$	$p_Z = 0.674$, $p_S = 0.004$	SUV_{95} remains predictive after adjustment, but mediation is not supported

Key point: C3 is strong, but C1 and C2 are not satisfied.

Results

Comparison across complementary methods

Method	Main result	Interpretation
Freedman proportion explained	$PE = -0.281$	Adding SUV_{95} did not attenuate the treatment effect in a way supporting surrogacy
Penalized logistic regression	same qualitative result	Main conclusions were robust to small-event shrinkage
Wang–Taylor sensitivity	$PE_{risk} \approx 0.476$	Some risk-pattern explanation, but not enough to establish surrogate validity
Causal mediation	no significant indirect effect	No evidence that SUV_{95} mediates treatment effect

Across methods: none provided convincing evidence that SUV_{95} is a validated surrogate endpoint.

Results

Predictive performance and model improvement

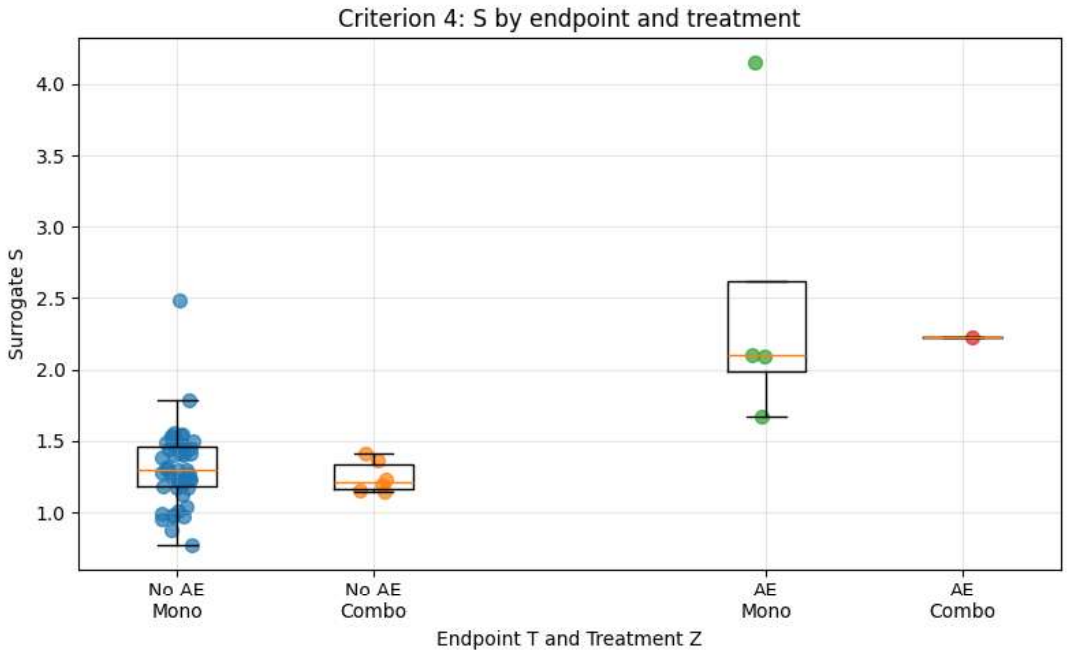
Model	AUC	Meaning
Treatment only ($T \sim Z$)	0.543	weak discrimination
SUV_{95} only ($T \sim S$)	0.981	excellent discrimination
Treatment + SUV_{95} ($T \sim Z + S$)	0.981	no meaningful gain beyond SUV_{95} alone

- Adding SUV_{95} markedly improved model fit
- Treatment alone contributed little predictive information

Interpretation: SUV_{95} is strongly prognostic, but prognostic value does not establish surrogacy.

Results

Criterion 4 visualization



SUV_{95} remains associated with AE risk, while treatment shows no clear residual contribution.

Results

Integrated interpretation

Pathway component	Evidence	Status
$Z \rightarrow S$	no	not supported
$Z \rightarrow T$	no	not supported
$S \rightarrow T$	yes	strongly supported
$Z \rightarrow S \rightarrow T$	no	surrogacy not established

Bottom line: SUV_{95} predicts toxicity risk, but surrogate validity was not established.

Conclusion

Overall conclusion

- Lung SUV_{95} strongly predicts lung adverse events
- Treatment did not significantly affect SUV_{95}
- Treatment also did not significantly affect the endpoint
- Therefore, the full surrogate pathway was not demonstrated in this dataset

SUV_{95} is prognostic, not a validated surrogate endpoint

Conclusion

Future work

- Evaluate alternative SUV biomarkers
- Examine regional or lobe-specific lung segmentation
- Test segmentation-related robustness more extensively
- Assess surrogate behavior in larger or multi-study datasets

Support: Freedman Proportion Explained

Coefficient-based measure

Step 1: Fit logistic models

C2: $\text{logit}(P(T=1)) = \beta_0 + \beta_1 Z \quad \Rightarrow \quad \beta_{\text{unadj}} = \beta_1$

C4: $\text{logit}(P(T=1)) = \delta_0 + \delta_1 Z + \delta_2 S \quad \Rightarrow \quad \beta_{\text{adj}} = \delta_1$

Proportion explained:

$$PE = 1 - \frac{\beta_{\text{adj}}}{\beta_{\text{unadj}}}$$

Interpretation:

- Reduction in treatment effect after adjusting for biomarker
- Works on coefficient (log-odds) scale

Support: Risk-Based Proportion Explained

Predicted probability measure

Model (C4):

$$\text{logit}(P(T = 1)) = \delta_0 + \delta_1 Z + \delta_2 S$$

Predicted risks:

$$P(T = 1 \mid Z, S) = \sigma(\delta_0 + \delta_1 Z + \delta_2 S)$$

$$AA = P(T = 1 \mid Z = 1,\ S = S_{\text{treated}})$$

$$AB = P(T = 1 \mid Z = 1,\ S = S_{\text{control}})$$

$$BB = P(T = 1 \mid Z = 0,\ S = S_{\text{control}})$$

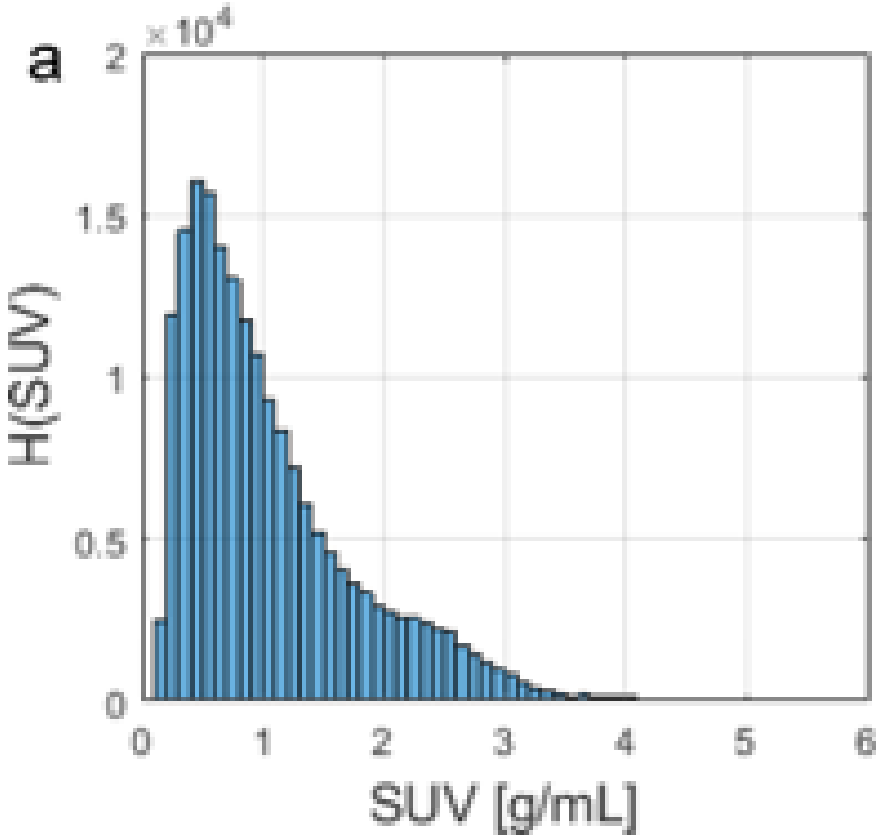
Risk-based proportion explained:

$$PE_{risk} = \frac{AA - AB}{AA - BB}$$

From single biomarker to structured surrogate analysis

- Different SUV variants
- Different Time Points

Metric	What it represents
SUV95	upper tail **(Meaningful)**
TailMean	smoothed upper tail **(Robustness)**

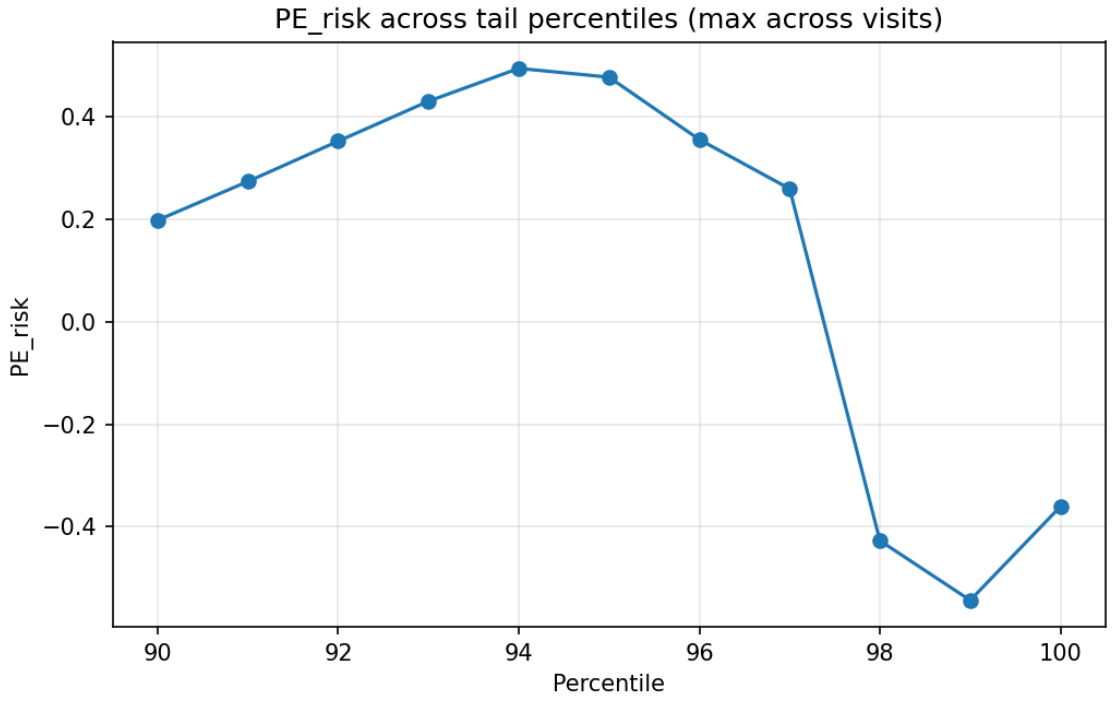


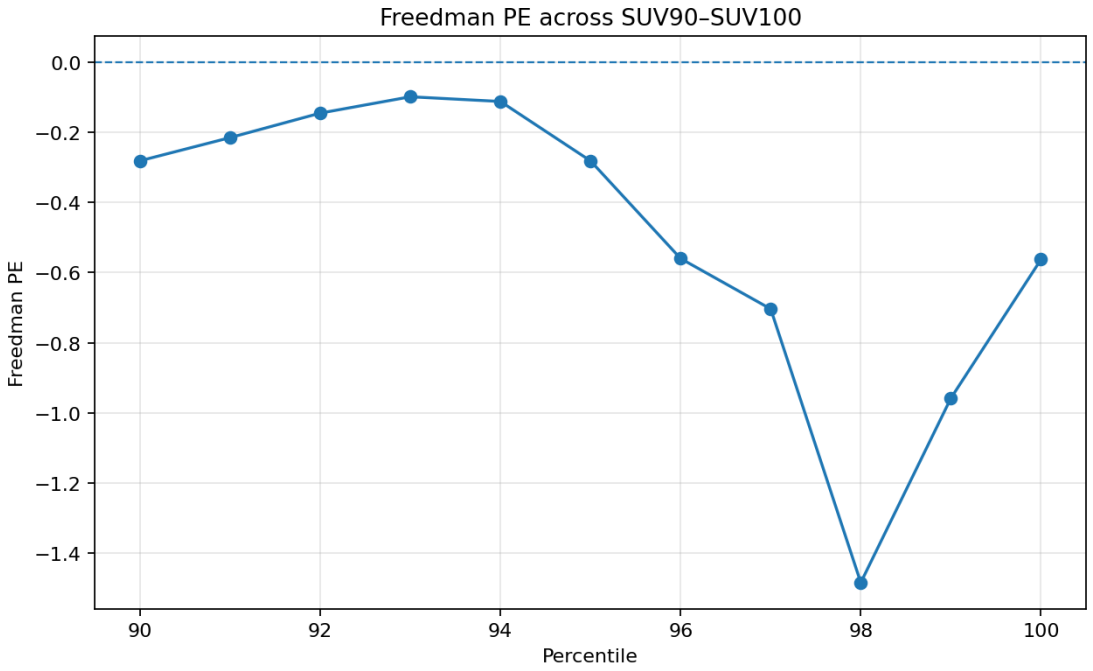
T Upper-Tail Biomarker Signal in Lung SUV

Biomarker	C1	C2	C3	C4
SUV95	✕	✕	Yes	✕
SUV96	✕	✕	Yes	✕
SUV97	✕	✕	Yes	✕
SUV98	✕	✕	Yes	✕
SUV99	✕	✕	✕	✕
SUVmax	✕	✕	✕	✕

med fiz

Upper-Tail Biomarker Signal in Lung SUV





Thank you
for your attention

med fiz