Confident and Logical Selection of the Cut-point of a Biomarker for Patient Targeting

Dr Yang Han

Department of Mathematics
The University of Manchester

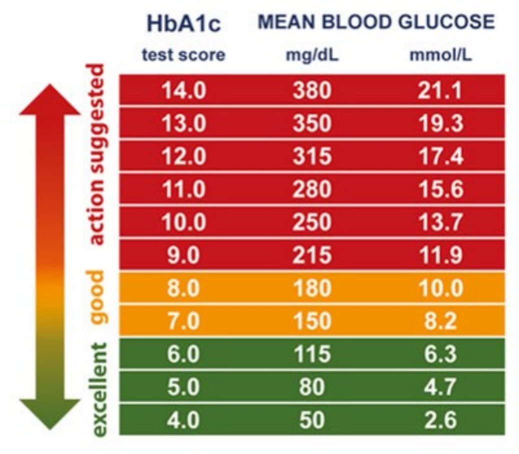
(Joint with Jason Hsu, Szu-Yu Tang, Hui-Min Lin)

Outline

- What is patient-targeting for a targeted therapy
- Why efficacy measure should respect logical relationships
- How to derive exact simultaneous confidence intervals
- Which approach to use

Patient-targeting for a targeted therapy

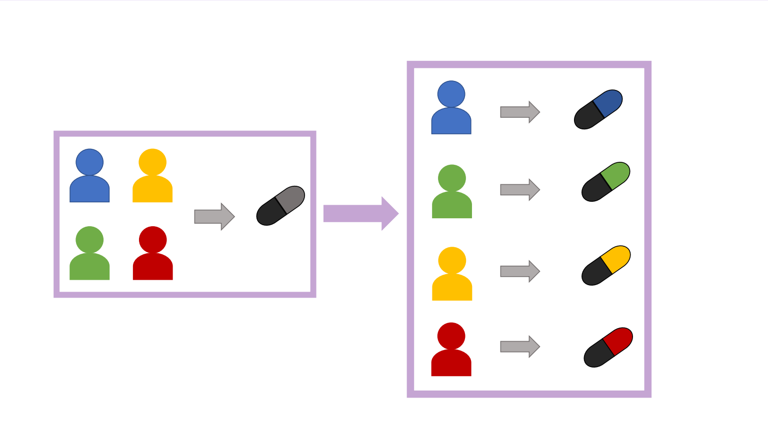
- Personalised (Precision) medicine aims to discover treatments that are more effective on average in one subgroup of patients than its complementary subgroup.



HbA1c test score	MEAN BLOOD GLUCOSE mg/dL	mmol/L
14.0	380	21.1
13.0	350	19.3
12.0	315	17.4
11.0	280	15.6
10.0	250	13.7
9.0	215	11.9
8.0	180	10.0
7.0	150	8.2
6.0	115	6.3
5.0	80	4.7
4.0	50	2.6

Patient-targeting for a targeted therapy

- Personalised medicine: a move away from a 'one size fits all' to one which uses new approaches to better manage patients' health and target therapies.



Patient-targeting for a targeted therapy

- For Type 2 Diabetes (T2DM), change in hemoglobin A1c (HbA1c) is the usual clinical measure of a treatment's effect.
- For Alzheimer's Disease (AD), change in Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) is a common measure of a treatment's effect.

Patient-targeting for a targeted therapy

Modelling for biomarker cut-point selection

Consider a two-arm randomized controlled trial (RCT). Denote 'treatment' and 'control' by Rx and C respectively,

$$Y_{ir_i} = \mu + \tau_i + \boldsymbol{\alpha} \boldsymbol{Z}_{r_i} + \beta X_{ir_i} + \gamma_i X_{ir_i} + \epsilon_{ir_i},$$

$$i = Rx \text{ or } C, \quad r_i = 1, \ldots, n_i$$

where

- τ is the treatment (Rx) vs. control (C) main effect
- $\boldsymbol{\alpha}$ represents covariate and block effects
- β is the biomarker's main effect, γ represents treatment $\times$ biomarker interaction, and ϵ_{ir_i} are iid Normal$(0, \sigma^2)$ with σ^2 unknown

Patient-targeting for a targeted therapy

Modelling for biomarker cut-point selection

Consider a two-arm randomized controlled trial (RCT). Denote ‘treatment’ and ‘control’ by *Rx* and *C* respectively,

$$Y_{ir_i} = \mu + \tau_i + \boldsymbol{\alpha Z}_{r_i} + \beta X_{ir_i} + \gamma_i X_{ir_i} + \epsilon_{ir_i},$$

$$i = Rx \text{ or } C, \quad r_i = 1, \ldots, n_i$$

- With targeted therapy, instead of using baseline for adjusting imbalance, we use baseline as a biomarker because sicker patients benefit more
- Having Region as a blocking factor allows inference on a common efficacy being made

Patient-targeting for a targeted therapy

Desirable properties for cut-point selection methods

- **Equipoise**: Provide inference on the marker-positive patients, and on the marker-negative patients, at all cut-point values
- **Subgroup logic-respecting**: Ensure inference respects natural relationships among the subgroups and all-comers
- **Cut-point logic-respecting**: Inferring clinically meaningful efficacy at a cut-point value implies meaningful efficacy at all higher cut-point values are inferred
- **Confidence**

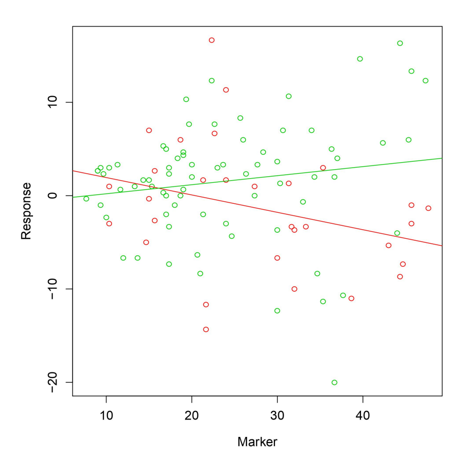
Patient-targeting for a targeted therapy

Motivating example

- Schnell et al. (2017) described an Alzheimer Disease study which compared three doses of a new treatment (doses 1, 2, and 3) with a negative control (dose 0, a placebo).
- Response in this study is *improvement* in ADAS-Cog11 (baseline ADAS-Cog11 minus final ADAS-Cog11) after 24 weeks.

Thresholding a continuous biomarker

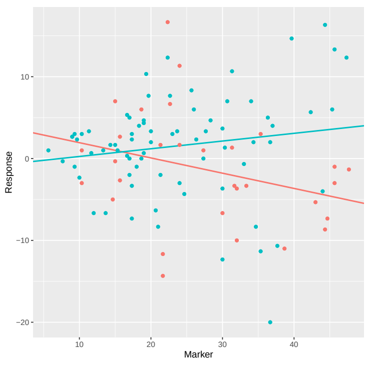
- For Alzheimer's Disease, change in ADAS-Cog is a measure of a treatment's effect.



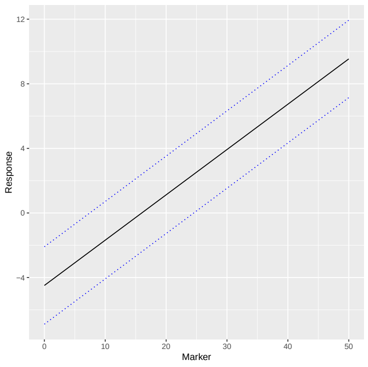
*Red represents 'control', green represents 'treatment'.

Thresholding a continuous biomarker

- For Alzheimer's Disease, change in ADAS-Cog is a measure of a treatment's effect.



(a) Improvement vs. Marker



(b) Efficacy vs. Marker

Patient-targeting for a targeted therapy

Suppose a biomarker has values in a range $[a, b]$, and each potential cut-point value $c \in (a, b)$ dichotomizes the patients into

- g_c^- subgroup with biomarker values $x \in [a, c)$
- g_c^+ subgroup with biomarker values $x \in [c, b]$

Two layers of multiplicity

- Multiplicity of subgroups for each cut-point value: assessing efficacy in g_c^-, g_c^+, and $\{g^-, g^+\}$ simultaneously
- Multiplicity of searching through infinitely many cut-points

Efficacy measure should respect logical relationships

With η denoting the efficacy of *Rx* vs. *C*, a clinically meaningful efficacy is achieved if $\eta > \delta$.

For a cut point *c*, denote the true (unknown) efficacy in g_c^+, g_c^-, and all-comers $\{all\}$ by $\eta_{g_c^+}, \eta_{g_c^-}, \eta_{\{all\}}$ respectively.

Definition of logic-respecting

An efficacy measure is **logic-respecting** if $\eta_{\{all\}} \in [\eta_{g_c^-}, \eta_{g_c^+}]\ \forall c \in (a, b)$.

Inference should respect natural relationships

Subgroup Sensitivity: Since $\eta_{g_c^+} > \delta$ and $\eta_{g_c^-} > \delta$ imply $\eta_{\{all\}} > \delta$ for any logic-respecting efficacy measure, the following requirement is desirable.

Definition

A method is *Subgroup Sensitive* if both $\eta_{g_c^+} > \delta$ and $\eta_{g_c^-} > \delta$ being inferred implies $\eta_{\{all\}} > \delta$ will be inferred.

Inference should respect natural relationships

Subgroup Specificity: Since $\eta_{\{all\}} > \delta$ implies either $\eta_{g_c^+} > \delta$ or $\eta_{g_c^-} > \delta$ or both for any logic-respecting efficacy measure, the following requirement may be desirable as well.

Definition

A method is *Subgroup Specific* if once $\eta_{\{all\}} > \delta$ is inferred, then at least one of $\eta_{g_c^+} > \delta$ and $\eta_{g_c^-} > \delta$ is inferred as well.

Principles for logical biomarker cut-point selection

- An efficacy measure is logic-respecting if $\eta_{\{all\}} \in [\eta_{g_c^-}, \eta_{g_c^+}], \ \forall c \in [a, b]$

Logical inference between subgroups and all-comers

- Subgroup sensitivity: $\{\eta_{g_c^-} > \delta\} \cap \{\eta_{g_c^+} > \delta\} \Rightarrow \{\eta_{all} > \delta\}$
- Subgroup specificity: $\{\eta_{all} > \delta\} \Rightarrow \{\eta_{g_c^-} > \delta\} \cup \{\eta_{g_c^+} > \delta\}$

Efficacy measure should respect logical relationships

- An efficacy measure η of *Rx* vs. *C* is a function of μ^{Rx} and μ^{C}.
- For any $[\delta_1, \delta_2] \subseteq [a, b]$, let $\pi_{[\delta_1,\delta_2]} = \int_{\delta_2}^{\delta_1} p(x)dx$, with $\pi_{[a,b]} = 1$.
- Within each of the *Rx* and *C* arms, the expected treatment effect in an interval $[\delta_1, \delta_2]$ can be calculated as

$$\mu^{Rx}_{[\delta_1,\delta_2]} = \int_{\delta_1}^{\delta_2} \mu^{Rx}(x)p(x)dx/\pi_{[\delta_1,\delta_2]}$$

$$\mu^{C}_{[\delta_1,\delta_2]} = \int_{\delta_1}^{\delta_2} \mu^{C}(x)p(x)dx/\pi_{[\delta_1,\delta_2]}$$

Efficacy measure should respect logical relationships

- True efficacy over the entire interval $[a, b]$ is

$$\eta_{[a,b]} = \eta\left(\int_a^b \mu^{Rx}(x)p(x)dx, \int_a^b \mu^{C}(x)p(x)dx\right)$$

- The efficacy within the intervals $[a, c)$ and $[c, b]$ are

$$\eta_{[a,c)} = \eta\left(\int_a^c \mu^{Rx}(x)p(x)dx/\pi_{[a,c)}, \int_a^c \mu^{C}(x)p(x)dx/\pi_{[a,c)}\right)$$
$$\eta_{[c,b]} = \eta\left(\int_c^b \mu^{Rx}(x)p(x)dx/\pi_{[c,b]}, \int_c^b \mu^{C}(x)p(x)dx/\pi_{[c,b]}\right)$$

- Combining them weighted by the prevalence $\pi_{[a,c]}$ and $\pi_{[c,b]}$ gives

$$\overline{\eta}_{[a,b]} = \pi_{[a,c)}\eta_{[a,c)} + \pi_{[c,b]}\eta_{[c,b]}$$

Efficacy measure should respect logical relationships

- However, in general, $\overline{\eta}_{[a,b]} \neq \eta_{[a,b]}$. The one example for which $\overline{\eta}_{[a,b]} = \eta_{[a,b]}$ is when efficacy is defined as a *difference*, $\eta = Rx - C$.
- The case of efficacy being a ratio, $\eta = Rx/C$, is such that one cannot take for granted that the efficacy measure is logic-respecting.

Example on odds ratio

Response rate	g^-	Half g^- + Half g^+	g^+
Rx	0.25	0.5	0.75
C	0.1	0.3	0.5
Odds Ratio	3	$2\frac{1}{3}$	3

Table: Using Odds Ratio as efficacy measure precludes Subgroup Sensitivity

Odds Ratio for the g^- and g^+ subgroups are both equal to 3, yet for the combined 50/50 mixture population the Odds Ratio is $2\frac{1}{3}$. Suppose the clinically meaningful efficacy threshold is 2.5. As sample size approaches infinity, any consistent method will declare efficacy in g^- and g^+ but declare a lack of efficacy in $\{g^-, g^+\}$, violating the Subgroup Sensitivity principle.

How to derive exact confidence intervals

Let $e(x)$ denote the point-wise efficacy of Rx vs. C.
For the linear model,

$$\begin{aligned}
\mu^{Rx}(x) &= \mu + \boldsymbol{\alpha Z} + \tau_{Rx} + (\beta + \gamma_{Rx})x \\
\mu^{C}(x) &= \mu + \boldsymbol{\alpha Z} + \tau_{C} + (\beta + \gamma_{C})x \\
e(x) = \mu^{Rx}(x) - \mu^{C}(x) &= (\tau_{Rx} - \tau_{C}) + (\gamma_{Rx} - \gamma_{C})x \\
&= \tau + \gamma x
\end{aligned}$$

$$P\{\hat{e}_L(x) < e(x) < \hat{e}^U(x) \;\; \forall x \in [a, b]\} = 1 - \alpha \qquad (1)$$

$$E_{\text{Band}} = \{\hat{e}_L(x) < e(x) < \hat{e}^U(x) \;\; \forall x \in [a, b]\}$$

Deriving confidence intervals for $\eta_{\{all\}}$, $\eta_{g_c^+}$, $\eta_{g_c^-}$

One can further deduce confidence intervals $I^{\pm}$, I_c^+, I_c^- for $\eta_{\{all\}}$, $\eta_{g_c^+}$, and $\eta_{g_c^-}$ for all $c \in [a, b]$

$$
\begin{aligned}
I_c^- &= \left(\frac{\int_a^c \hat{e}_L(x)p(x)dx}{\int_a^c p(x)dx}, \frac{\int_a^c \hat{e}^U(x)p(x)dx}{\int_a^c p(x)dx} \right) \\
I_c^+ &= \left(\frac{\int_c^b \hat{e}_L(x)p(x)dx}{\int_c^b p(x)dx}, \frac{\int_c^b \hat{e}^U(x)p(x)dx}{\int_c^b p(x)dx} \right) \qquad (2) \\
I^{\pm} &= \left(\frac{\int_a^b \hat{e}_L(x)p(x)dx}{\int_a^b p(x)dx}, \frac{\int_a^b \hat{e}^U(x)p(x)dx}{\int_a^b p(x)dx} \right)
\end{aligned}
$$

Deriving confidence intervals for $\eta_{\{all\}}$, $\eta_{g_c^+}$, $\eta_{g_c^-}$

$$E_{\text{CI}} =$$

$$\left\{ \frac{\int_a^c \hat{e}_L(x)p(x)dx}{\int_a^c p(x)dx} < \frac{\int_a^c e(x)p(x)dx}{\int_a^c p(x)dx} < \frac{\int_a^c \hat{e}^U(x)p(x)dx}{\int_a^c p(x)dx} \text{ and} \right.$$

$$\frac{\int_c^b \hat{e}_L(x)p(x)dx}{\int_c^b p(x)dx} < \frac{\int_c^b e(x)p(x)dx}{\int_c^b p(x)dx} < \frac{\int_c^b \hat{e}^U(x)p(x)dx}{\int_c^b p(x)dx} \text{ and}$$

$$\frac{\int_a^b \hat{e}_L(x)p(x)dx}{\int_a^b p(x)dx} < \frac{\int_a^b e(x)p(x)dx}{\int_a^b p(x)dx} < \frac{\int_a^b \hat{e}^U(x)p(x)dx}{\int_a^b p(x)dx},$$

$$\left. \forall c \in [a, b] \right\}.$$

Deriving confidence intervals for $\eta_{\{all\}}$, $\eta_{g_c^+}$, $\eta_{g_c^-}$

Theorem

Simultaneous confidence intervals (1) and (2) together have exact (not conservative) coverage probability $1-\alpha$.

$$Pr\{E_{\text{Band}} \cap E_{\text{CI}}\} = 1 - \alpha$$

Among methods based on confidence bands, the one based on *constant width* confidence band achieves all the inferential requirements because then widths of the confidence intervals for $\eta_{g_c^+}$, $\eta_{g_c^-}$, and $\eta_{\{all\}}$ remain equal, for all c.

Two layers of multiplicity

- Multiplicity of subgroups for each cut-point value: assessing efficacy in g_c^-, g_c^+, and $\{all\}$ simultaneously
- Multiplicity of searching through infinitely many cut-points

Our **Simultaneous CIs** based method guarantees

$$\inf_{c \in [a,b]} P\{\eta_{g_c^-} \in I_c^- \text{ and } \eta_{g_c^+} \in I_c^+ \text{ and } \eta_{\{all\}} \in I^\pm\} \geq 1 - \alpha.$$

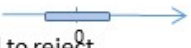
It controls the FWER of testing the infinitely many 1-sided null hypotheses below, six null hypotheses for each c:

$$\begin{aligned} H_{\leq}^{+} : \eta_{g_c^+} \leq \delta, &\quad H_{\geq}^{+} : \eta_{g_c^+} \geq \delta \\ H_{\leq}^{-} : \eta_{g_c^-} \leq \delta, &\quad H_{\geq}^{-} : \eta_{g_c^-} \geq \delta \\ H_{\leq}^{\pm} : \eta_{\{all\}} \leq \delta, &\quad H_{\geq}^{\pm} : \eta_{\{all\}} \geq \delta. \end{aligned}$$

Which approach to use

Why Confidence Intervals? not p-values?

Making decision based on CI controls directional error rate

- From a *family* of tests to CI
  - Make a separate 5% test for every θ
  - Observe *Data Y*
  - Put all θ not rejected into a set *C*
  - *C* is a 95% Confidence Set for θ
- From CI to directional tests
  - If CI cover 0 
    - then fail to reject
  - If CI entirely > 0 
    - then infer > 0
  - If CI entirely < 0 
    - then infer < 0

Control of directional error rate

No (with logical relationships)

H_0: $\theta_+ = 0$

H_0: $\theta_- = 0$

H_0: $\theta_\pm = 0$

Yes

H_0: $\theta_+ \le 0$	H_0: $\theta_+ \ge 0$
H_0: $\theta_- \le 0$	H_0: $\theta_- \ge 0$
H_0: $\theta_\pm \le 0$	H_0: $\theta_\pm \ge 0$

Shiny App

Simultaneous Confidence Intervals for Efficacy

Upload Data Set (csv)

Browse... SchnellADdata - Copy.csv

Upload complete

Cut-point multiplicity adjustment

Simultaneously for All Cut-point Values

Alpha (in percents)

1 10 50

1 6 11 16 21 26 31 36 41 46 50

Proportion of marker negative subjects in population (in percents)

0 50 100

0 10 20 30 40 50 60 70 80 90 100

Treatment Variable

treatment

Control Value

0

Biomarker Variable

severity

Data filter cut-point c0 (subjects w/ value < C0 are excluded)

5.667 22.5

5.667 7.5 9.3 11.1 12.9 14.7 16.5 18.3 20.1 21.922.5

Marker +/- cut-point c1 (subjects w/ value < C1 are marker -)

10.333 22.35 45.667

Main | Help

Estimated mean outcome

	Negative	Positive	Mixture
Rx	1.051	2.098	1.575
C	-0.167	-2.204	-1.185

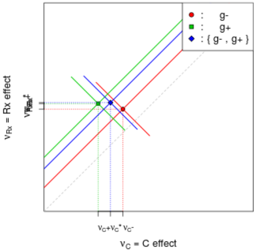
Estimated difference of means

	Negative	Positive	Mixture
Rx-C	1.954	4.444	3.506

Simultaneous confidence intervals (adjusted for the multiplicity of all cut-point values)

	Negative	Positive	Mixture
Upper	6.377	8.868	7.929
Lower	-2.470	0.021	-0.917

Download Tables

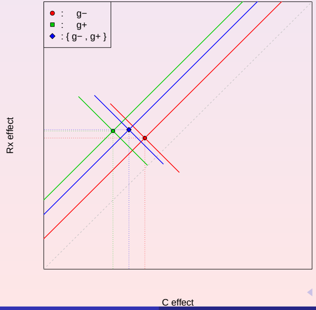


Download Plot

Example

We analysed the Alzheimer data at $\alpha = .10$, setting $\delta = 0$ for illustration, $\overline{c_M^+} = 22.34$.

	η_{g^-}	$\eta_{\{all\}}$	η_{g^+}
Upper bound	6.376	7.929	8.867
Estimate	1.953	3.506	4.443
Lower bound	−2.471	−0.917	0.020



Summary

- For continuous outcome modelled linearly and efficacy defined as a difference of means, the confidence band method in this work achieves all the logical inference requirements.
- For binary and time-to-event outcomes and logic-respecting efficacy measures such as Relative Response and Ratio of Medians, corresponding development requires non-trivial application of the Subgroup Mixable Estimation principle.

Moving forward to individualized medicine

Individualized medicine is a focus of our current research.

The question is: “Is it true that at least 80% of the patients will benefit more from being given *Rx* than from being given *C* by at least an amount δ?”

Construction of simultaneous tolerance bands

Over the covariate range $x \in [a, b]$,

- **Simultaneous Tolerance Bands (STBs)**

  $$P_{\mathcal{E}}\left\{ P_{y_x}\left\{ L_s(x;\mathcal{E}) < y_x < U_s(x;\mathcal{E}) | \mathcal{E} \right\} \geq \beta \ \text{ for all } \ x \in [a, b] \right\} = 1-\alpha.$$

- **Weighted Simultaneous Tolerance Bands (WSTBs)** with distribution $x \sim F(x)$

  $$P_{\mathcal{E}}\left\{ \int_a^b P_{y_x}\left\{ L_w(x;\mathcal{E}) \leq y_x \leq U_w(x;\mathcal{E}) | \mathcal{E} \right\} dF(x) \geq \beta \right\} = 1-\alpha.$$

References

Han, Y., Tang, S., Lin, H. and Hsu, J. (2022).
Exact simultaneous confidence intervals for logical selection of a biomarker cut-point.
Biometrical Journal. **64**(2), 272-289.

Han, Y., Sun, Y., Wang, L., Liu, W. and Bretz, F. (2023).
Statistical calibration for infinite many future values in linear regression: simultaneous or pointwise tolerance intervals or what else?
Journal of the Royal Statistical Society (C). **65**(4), 641-648

Han, Y., Wang, L., Liu, W. and Bretz, F. (2025).
Multiple-use calibration for all future values and exact two-sided simultaneous tolerance intervals.
Journal of Computational and Graphical Statistics. 34(1), 155-161.