Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

When the wild type hits the ceiling: identifiability of mutation–lineage interactions in pharmacogenomic panels

Jun Bao
Southeast University
Tianyu Wang
Erasmus University Rotterdam
Chao Song
230219371@seu.edu.cn

Southeast University

Research Article

Keywords: censoring, IC50, pharmacogenomics, mutation–lineage interaction, GDSC, CCLE, partial identification, estimator selection, targeted therapy

Posted Date: October 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11068462/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

When the wild type hits the ceiling: identifiability of mutation–lineage interactions in pharmacogenomic panels

Jun Bao[1,2,†], Tianyu Wang[3,†], Chao Song[1,2,*]

[1] State Key Laboratory of Digital Medical Engineering, Southeast University, Nanjing 210096, China

[2] School of Biological Science & Medical Engineering, Southeast University, Nanjing 210096, China

[3] Department of Gastroenterology and Hepatology, Erasmus MC, University Medical Center, Rotterdam 3015 GD, the Netherlands

* Correspondence: 230219371@seu.edu.cn (C.S.)

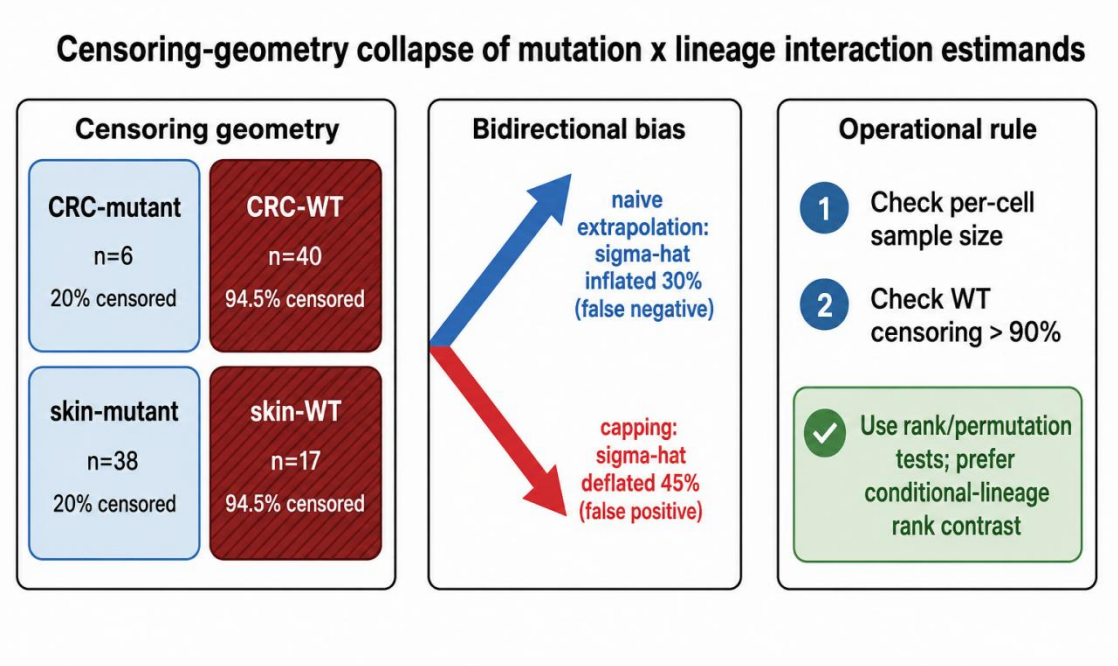
† These authors contributed equally to this work.

Email addresses:

JB: kaijun@seu.edu.cn

TW: t.w.wang@erasmusmc.nl

CS: 230219371@seu.edu.cn



Graphical Abstract

Abstract

Background: Targeted cancer drugs often work against a mutation in one tissue but not another: BRAF V600E inhibitors cure melanoma yet fail colorectal cancer. Studying such mutation-by-tissue interactions relies on public cell-line panels (GDSC, CCLE), where the wild-type lines are, by construction, insensitive to the drug, so 94–95% of them are recorded at the assay ceiling (right-censored).

Results: This censoring geometry, not the choice of estimator, is what makes the mutation × lineage interaction effect practically unidentifiable: extrapolating the ceiling and capping at it bias the result in opposite, predictable directions, and the sharp interval does not shrink with sampling under any positive censoring, and beyond ~90% censoring even the parametric minimum detectable effect exceeds practical magnitudes. Surveying GDSC, we find a three-level bottleneck: 94.5% of units lack a hotspot mutation, 45% of the remainder are too small, and 36.4% of the survivors are too heavily censored to inform the interaction.

Conclusions: We provide a sign-reversal condition, a minimum-detectable-effect and cell-level admission rule, and an operational two-level gate: check sample size, then censoring; when censoring is heavy, compare the mutant cells alone rather than the full interaction.

Keywords: censoring; IC50; pharmacogenomics; mutation–lineage interaction; GDSC; CCLE; partial identification; estimator selection; targeted therapy

Background

A targeted drug can be highly effective against a mutation in one tissue yet nearly useless in another: the BRAF inhibitor vemurafenib achieves ~53% response in BRAF V600E melanoma but only ~5% in BRAF V600E colorectal cancer, a discrepancy driven by tissue-specific EGFR feedback reactivation [18,19]. Studying such mutation×tissue interactions relies on public pharmacogenomic panels, whose wild-type cell lines are—by construction—insensitive to the drug and therefore censored at the assay ceiling. This paper shows that this geometry, not the analyst's estimator, is what makes the interaction effect unidentifiable.

Large pharmacogenomic panels — the Genomics of Drug Sensitivity in Cancer (GDSC) and the Cancer Cell Line Encyclopedia (CCLE) — underpin the discovery of predictive biomarkers: which genomic alterations predict sensitivity to which drugs. The dominant statistical framework is a *marginal* association test between a drug-response metric and a single genomic feature, estimated by MANOVA or elastic-net models [1,2].

A long-standing and well-characterised complication is that dose-response metrics are **censored**. Reported IC50 values are routinely extrapolated beyond the maximum screened concentration, and it is now established that activity-area (AUC/AAC) metrics are

more robust across datasets than IC50 [3,4,7,8,5,10], a conclusion reinforced by systematic benchmarking [14,17], and that GR metrics address a separate confound of division rate [6]. Censoring-aware curve fitting [5] and interval-censored regression have been proposed for the *marginal* drug–gene association [16].

Here we make a change of statistical object. Tissue-of-origin is a first-order determinant of drug response, yet it is often treated as a nuisance covariate or ignored [2,9]. When a mutation's effect is expected to vary by lineage, the estimand of interest is the **mutation × lineage interaction** — a 2×2 contrast across four cells (two lineages × mutant/WT). The marginal association requires information in a single cell; the interaction requires information in **all four**.

We show that the interaction estimand becomes practically non-informative under a geometry that is common for selectively targeted drugs: the two WT cells are near-totally right-censored (94–95%) because WT lines are insensitive to the drug, while the two mutant cells carry the information. In this geometry the 2×2 interaction collapses to a conditional-lineage contrast, and the two standard workarounds fail in opposite, predictable directions — naive extrapolation inflates the residual SD by ~30% (false negatives), while capping inflates precision (false positives). We characterize the identifiability boundary analytically (Proposition 1), derive the minimum-detectable-effect and a cell-level admission heuristic (Proposition 2, Corollary 1), and quantify the prevalence of the failure in GDSC1/GDSC2 (a double bottleneck of per-cell sample size and censoring). We close with an operational estimator-selection rule.

The prior literature treats the robustness of *marginal* drug–gene associations; to our knowledge, the identifiability of the *interaction* estimand under per-cell censoring imbalance — and the resulting bidirectional, sign-reversible bias — has not been characterised. We frame the results as a reusable set of methodological criteria: Propositions 1–2 and Corollary 1 characterize the identifiability boundary, and rules R0–R5 translate them into a design checklist that an analyst can apply before running any mutation × lineage interaction test.

Methods

Data

We used the Genomics of Drug Sensitivity in Cancer (GDSC) fitted dose–response datasets, release 27 Oct 2023, for both GDSC1 (970 cell lines, 366 drugs) and GDSC2 (969 cell lines, 286 drugs). For each cell-line × drug pair we retained the natural-log IC50 (`LN_IC50`), the maximum screened concentration (`MAX_CONC`), the activity area (`AUC`), and the annotated putative target (`PUTATIVE_TARGET`). Cell-line metadata (Sanger model ID → DepMap ID → lineage) were taken from the Cancer Cell Line Encyclopedia (CCLE) `sample_info` table.

Somatic mutation calls were taken from `CCLE_mutations.csv` [13] (1,231,281 variant records across 19,537 genes). We restricted to non-silent variants (Missense, Nonsense,

Frameshift, In-frame indel, Splice site). Recurrently mutated "hotspot" variants were identified using the CCLE-provided `isTCGAhotspot` and `isCOSMIChotspot` flags (2.9% of non-silent calls).

Censoring definition

We defined a drug response as **right-censored** whenever the reported IC50 exceeded the maximum screened concentration (MAX_CONC),

$$\text{censored} \Leftrightarrow \text{LN_IC50} \geq \ln(\text{MAX_CONC}).$$

The per-cell censoring rate is $\pi_l = P(Y_l^* > c_l)$, where $c_l = \ln(\text{MAX_CONC}_l)$ and Y^* is the latent log-IC50. The observed response is $Y_l = \min(Y_l^*, c_l)$ with censoring indicator $d_l = \mathbf{1}\{Y_l^* > c_l\}$.

Estimators

Throughout, a "cell" denotes a group of cell lines (a stratum), not a single cell line. For a 2×2 mutation×lineage contrast with cells $l \in \{AM, AW, BM, BW\}$ (A/B = two lineages, M/W = mutant/wild-type), we compared six estimation strategies:

1. **naive OLS** — linear model on the extrapolated `LN_IC50`, with the interaction coefficient β_{int} and HC0/HC2/HC3 sandwich standard errors;
2. **capped OLS** — linear model on $\min(\text{LN_IC50}, \ln(\text{MAX_CONC}))$;
3. **bounded-IC50 transform** — the inhibition fraction at the maximum concentration under a fixed Hill slope, $1/(1 + (\exp(c)/\text{IC50})^{0.5})$ (a deterministic monotone transform of IC50, hence rank-equivalent to IC50), fitted with the same linear model;
4. **mutant-only rank** (primary, E1) — Mann–Whitney U test comparing the two mutant cells only (conditional lineage comparison), which discards the censored WT cells entirely; note the U test targets stochastic ordering, which coincides with a mean difference δ_M only under a location-shift model;
5. **Tobit** [12] — interval-censored Gaussian regression (reported in the Appendix);
6. **censoring-probability interaction (E6)** — the 2×2 interaction on the censoring indicator d_{il} (risk difference $(\pi_{AM} - \pi_{AW}) - (\pi_{BM} - \pi_{BW})$, with a delta-method standard error), which is fully identified without extrapolation and retains the 2×2 structure.

Welch–Satterthwaite effective degrees of freedom and permutation nulls were applied as specified. The mutant-only rank test is the primary endpoint because it discards the censored WT cells and is invariant to *WT-cell* censoring (though not to mutant-cell censoring, which creates ties; its power still depends on the mutant-cell n).

Simulation framework (A1)

We generated latent log-IC50 as $Y_{il}^* \sim \mathcal{N}(\mu_l, \sigma_l^2)$ in four cells with $n = (6,40,38,17)$ matching the BRAF V600E example, $\sigma_{\text{mut}} = 1.5$, $\sigma_{\text{WT}} = 0.8$, threshold $c = \ln 10$. The naive

extrapolated value for a censored observation was drawn as $Y_{il}^{*} + \epsilon$ with $\epsilon \sim \mathcal{N}(0,4)$ (SD = 2, reflecting the typical magnitude of GDSC's IC50 extrapolation beyond MAX_CONC); σ_{true} denotes the generative latent SD σ_l (1.5 for mutant, 0.8 for WT), so $\hat{\sigma}/\sigma_{\text{true}}$ compares the estimator's residual SD against the generative value. Cell means were calibrated to fixed censoring rates via $\mu_l = c + \sigma_l\, \Phi^{-1}(\pi_l)$: mutant cells at 20% censoring, WT cells at 94.5%. The interaction estimand is $\beta_{\text{int}} = (\mu_{AM} - \mu_{AW}) - (\mu_{BM} - \mu_{BW})$; the alternative sets a latent potency shift of 1.8 log units ($\approx$ 6-fold). We report empirical false-positive rate (FPR), size-adjusted power (5% quantile of each estimator's null), and the residual-SD ratio $\hat{\sigma}/\sigma_{\text{true}}$.

The two-dimensional grid (see the simulation grid below) varies the WT censoring rate $p \in \{0.2, 0.5, 0.8, 0.9, 0.945\}$ against the balanced cell counts $n \in \{(6,40,38,17),\ (20,20,20,20),\ (40,40,40,40)\}$ to separate the censoring-geometry effect from the small-cell ($n = 6$) effect.

Analytic propositions

Two propositions and one corollary (proofs in the Appendix) characterise the identifiability boundary:

- **Proposition 1 (partial identification of the estimand).** *(i) Non-parametric sharp bounds*. Under Type-I right-censoring with no distributional assumption beyond a finite support bound B, the sharp identified set for $\mu_l = E[Y_l^{*}]$ is $[\theta_l, \theta_l + \pi_l B]$ with $\theta_l = E[Y \wedge c]$; the interaction estimand therefore has a sharp interval of width $W = B \sum_l \pi_l$ that does not shrink with n, so no consistent estimator exists without extrapolation. *(ii) Parametric point identification*. Under a normal location-scale model with correctly specified σ, the MLE identifies μ_l whenever $\pi_l < 1$, so the interaction is estimable in sign; the only limit is information-driven (the MDE grows as censoring rises), not structural. *(iii) Parametric pseudo-limit*. The capped OLS estimand converges to the pseudo-limit $\tilde{\beta}_{\text{int}} = \sum_l s_l\ (c - \sigma\Lambda(\zeta_l))$, determined by (c, π, σ) rather than by the WT-cell means; this pseudo-limit can reverse the sign of the true interaction and bounds neither the true interaction nor its sign.
- **Proposition 2 (information bound).** Under the normal model the per-cell Fisher information is $I_1(\pi_l) = h(\pi_l)/\sigma^2$ with $h(\pi) = \Phi(\zeta) - \zeta\phi(\zeta) + \phi(\zeta)^2/(1 - \Phi(\zeta))$, $h(0) = 1$, $h(0.90) = 0.36$, $h(0.95) = 0.23$. The minimum detectable interaction effect is $\text{MDE} = \kappa\sigma\, (\sum_l [\, n_l h(\pi_l)]^{-1})^{1/2}$, $\kappa \approx 2.80$.
- **Corollary 1 (cell-level admission heuristic).** Replacing $h(\pi_l)$ by the un-censored fraction $1 - \pi_l$ is a *heuristic* that underestimates the information (because $h(\pi) > 1 - \pi$ for $\pi > 0$, with equality at $\pi = 0$), hence conservative for the MDE, giving an effective n dominated by the smallest $m_l = n_l(1 - \pi_l)$. As a planning rule, interaction inference is practical only when $\sum_l m_l^{-1} \le (\Delta^{*}/\kappa\sigma)^2$; a sufficient per-cell threshold is $m_l \ge 4(\kappa\sigma/\Delta^{*})^2$. Inverting, $\pi_W = 0.93$ requires $n_W \ge 460$, whereas panels have $n_W \le 50$, yielding MDE ≥ 1.5 log units — such a 2×2 is unlikely to support interaction inference. A sufficient condition for sign reversal of the capped OLS pseudo-limit is $\delta_M \delta_W > 0$ and $(1 - \pi_{W,\min})/(1 - \pi_{M,\max}) < |\delta_M/\delta_W| < 1$.

A1b survey of censoring geometry

We enumerated drug × gene × lineage units for targeted agents (PUTATIVE_TARGET resolving to a gene symbol) and, as a negative contrast, cytotoxic agents (mechanistic annotation such as “DNA crosslinker”, “microtubule stabiliser”, “antimetabolite”). For each unit we computed the mutant-cell and WT-cell counts and censoring rates, classifying a unit as *insufficient* when either cell had fewer than $n_{\min} = 5$ lines, and as *non-informative* when the WT censoring rate exceeded $c^* = 0.90$ (sensitivity at 0.80 and 0.95). The non-informative fraction of units among sufficient units was stratified by drug class. Note that this survey measures the two-cell mutant-vs-WT pair within a single lineage, not the four-cell mutation×lineage interaction; the fraction of non-informative four-cell interactions is not directly enumerated here.

Panoramic inference: positive and negative controls

Seven point-mutation positive controls (drug × gene × lineage) with established biology were tested within-lineage (mutant vs WT), with recall defined as correct direction + one-sided $p < 0.05$ + an effect-size floor ($|\beta| \geq 0.5$ for OLS, $|\delta| \geq 0.3$ for the rank test): dabrafenib/BRAF V600E, trametinib/BRAF V600E, nutlin-3a/TP53 (direction reversed: WT-sensitive), afatinib/EGFR L858R+exon19del, trametinib/NRAS Q61, alpelisib/PIK3CA H1047R/E545K, trametinib/KRAS G12/13. Negative controls were 806 randomly paired unrelated drug × gene × lineage units, used to calibrate each estimator’s empirical false-positive rate.

Results

Mechanism: bidirectional bias, invariant to sample size

σ̂ bias is a censoring-geometry effect, essentially invariant to n. In the A1 simulation (four cells $n = (6,40,38,17)$, WT censoring 94.5%), the naive extrapolated estimator inflated the residual SD by ~30% ($\hat{\sigma}/\sigma = 1.30$) while the capped estimator deflated it by ~45% ($\hat{\sigma}/\sigma = 0.55$), with the interaction coefficient itself essentially unbiased (bias < 0.004). The two-dimensional grid showed this ratio is essentially invariant to cell counts (Figure 1):

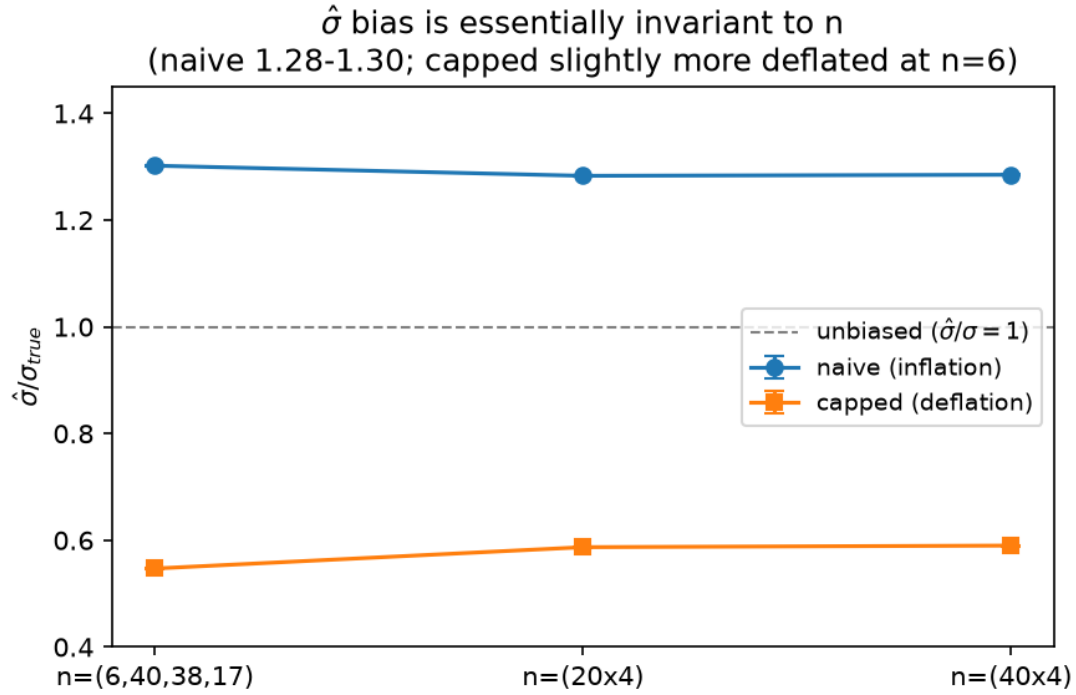


Figure 1

$\hat{\sigma}/\sigma_{\text{true}}$	n=(6,40,38,17)	n=(20×4)	n=(40×4)
naive (inflation)	1.302 ± 0.002	1.283 ± 0.002	1.285 ± 0.001
capped (deflation)	0.547 ± 0.001	0.587 ± 0.001	0.590 ± 0.001

The naive inflation is essentially flat across n (1.28–1.30). The capped deflation is slightly stronger at $n = 6$ (0.55 vs 0.59, a ~7% additional small-sample deflation superimposed on the censoring-geometry effect). The $\hat{\sigma}$ bias is therefore dominated by the censoring geometry, with a modest small-cell contribution for capping.

Inference failure is sample-size dependent. By contrast, the *visible* failure — low power and inflated FPR — depended on n. At 94.5% WT censoring, naive power rose from 0.32 (n=6) to 0.50 (n=20) to 0.81 (n=40), and the capped estimator's FPR inflation (0.076 at n=6) vanished at n=40 (0.050). These are two orthogonal results: the bias is structural and cannot be removed by enlarging n; the visibility of the failure is sample-size dependent. Even at n=40, naive power (0.81) remained below capped/bounded-IC50/rank (~0.99), with the 29% $\hat{\sigma}$ inflation still costing ~18 percentage points of power.

Estimator ranking (Table 1, N=6000–8000). (the replicate count differs across estimators because non-convergent Tobit fits were discarded) The mutant-only rank test was optimal for its target (the conditional-lineage contrast δ_M), and Satterthwaite degrees of freedom calibrated naive to FPR 0.051; because the rank and OLS estimators target different estimands, their powers are not directly comparable. Tobit showed 17–19% non-convergence and interaction-coefficient bias up to $+0.86$ (in the four-cell design with small

WT cells) that persisted under a homoscedastic DGP — a finite-sample boundary event whose sign and magnitude depend on the effective WT cell size m_W (see the Tobit boundary analysis in the Appendix), consistent with the per-cell admission heuristic (Discussion).

Estimator	FPR (H0)	size-adj. power	$\hat{\sigma}/\sigma$
naive + Satterthwaite	0.051	0.32	1.30
capped + HC3	0.076	0.66	0.55
bounded-IC50 + HC3	0.068	0.59	—
mutant-only rank	**0.045**	**0.73**	—
Tobit	0.074	0.05	conv. fail 17–19%

Table 1. Operating characteristics of the five OLS/rank/Tobit estimators (A1 simulation, WT censoring 94.5%, N=6000–8000). The mutant-only rank test is optimal for its target (δ_M, FPR 0.045, size-adjusted power 0.73); the OLS estimators target $\delta_M - \delta_W$ and are not directly comparable. Note that the estimators target different estimands: the mutant-only rank test targets δ_M (the conditional-lineage contrast), while the OLS estimators target $\delta_M - \delta_W$ (the interaction). Across $\delta_W \in \{0, 0.5, 1, 1.5\}$ the rank test's power was flat (0.71–0.74) while the OLS interaction power fell with the true interaction (see the estimator-target comparison in the Appendix); their powers are therefore not directly comparable.

Negative-control calibration. Across 806 unrelated drug × gene × lineage pairs, all three estimators were calibrated under *non-differential* censoring: naive FPR 4.8%, capped 3.0%, rank 5.3% (nominal 5%; Table 2). The naive estimator therefore does not produce spurious positives when censoring is unrelated to the mutation; its distortion is confined to *differential* censoring (see the marginal readout below).

Table 2. Negative-control false-positive rates (806 unrelated drug × gene × lineage pairs; nominal 5%).

Estimator	FPR
naive	4.8%
capped	3.0%
MWW rank	5.3%

Case dissection: BRAF V600E, CRC vs melanoma

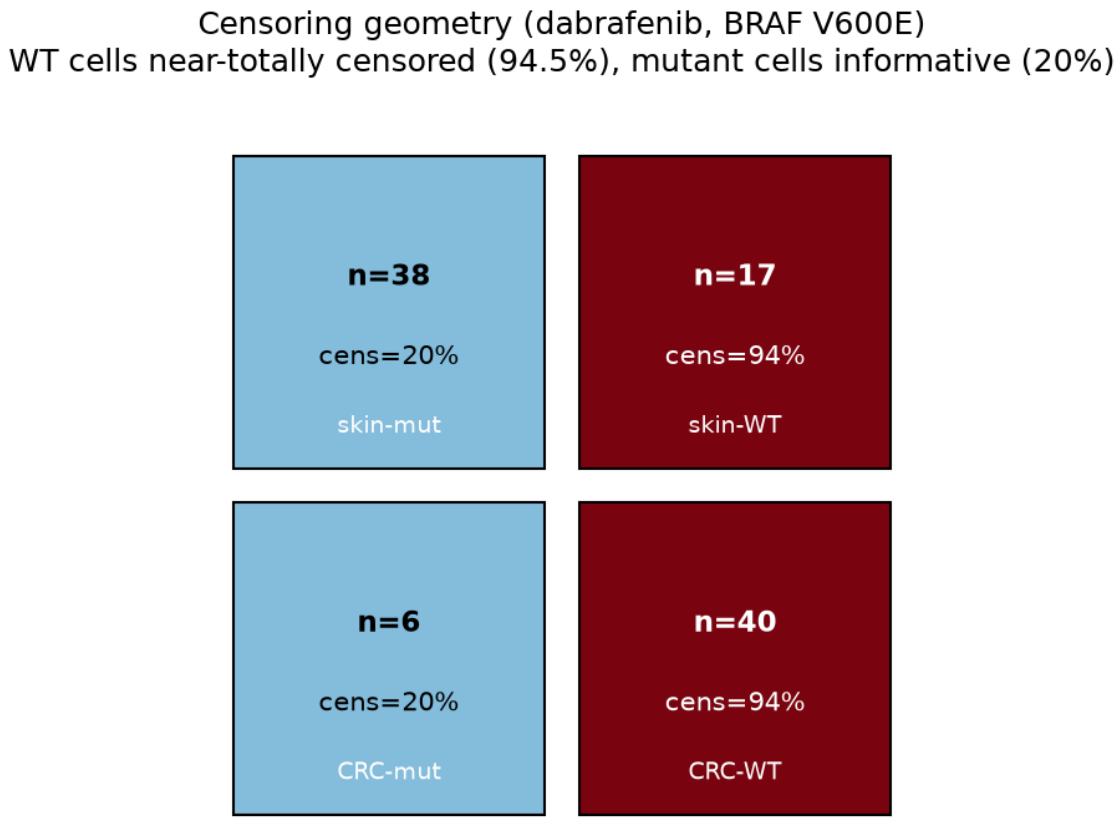


Figure 2

(Figure 2) For dabrafenib, the four cells were $n = (6,40,38,17)$ (CRC-mut, CRC-WT, skin-mut, skin-WT) with WT censoring 94–95%. The five OLS/rank/Tobit estimators (E6 reported separately in the Appendix) returned a stable interaction coefficient (β 1.6–2.25) but p-values spanning 0.41 to 0.0000, reproducing the simulated mechanism exactly: the coefficient is estimable in sign but the standard error — and hence the inference — is driven by the censoring geometry. This example is post-hoc and exploratory; the confirmatory layer is the panoramic identifiability survey (see the panoramic survey below).

Panoramic scope: the three-level bottleneck

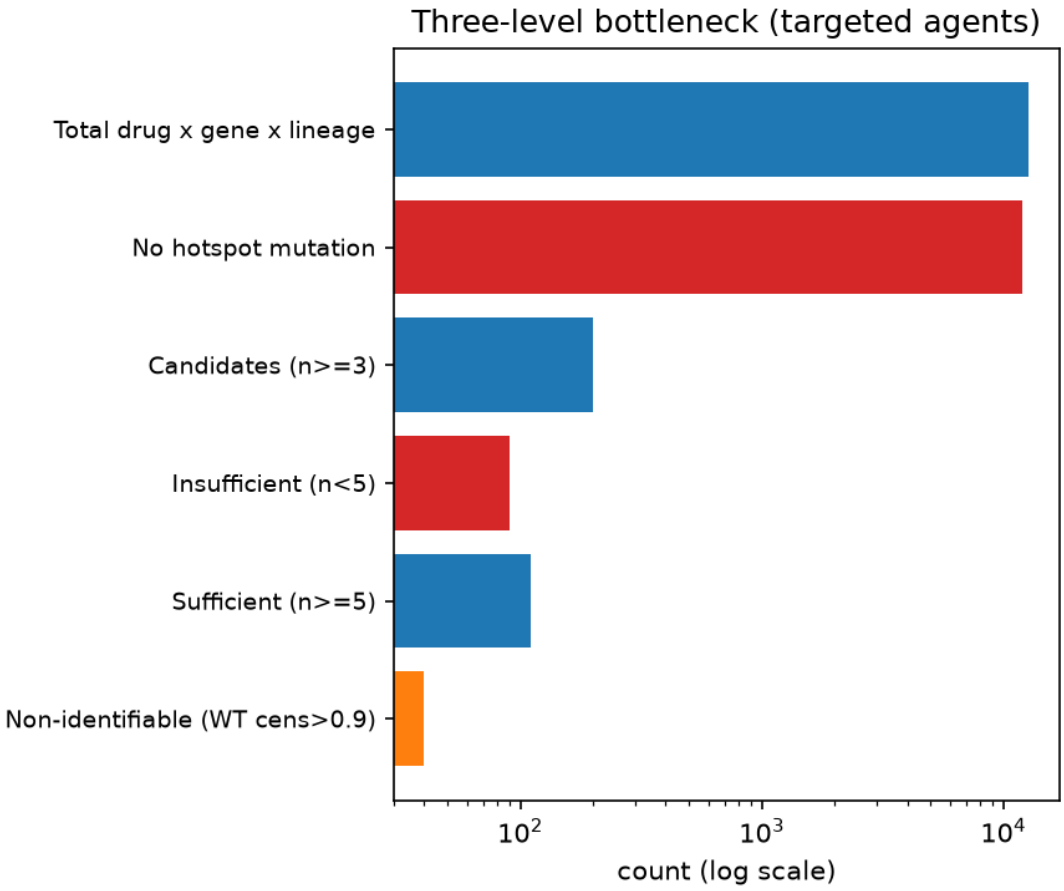


Figure 3

(Figure 3) **First level — no hotspot mutation.** Across 228 targeted agents, 94.5% of drug × target-gene × lineage units carried no hotspot mutation in the queried lineage, and are never tested in practice.

Second level — sample-size insufficiency. Of the 200 remaining candidates (mutant and WT cells each with ≥ 3 lines), 45.0% were insufficient (a cell with < 5 lines; Table 4).

Third level — censoring non-informativeness (the non-informative fraction). Of the 110 sufficient units, 36.4% had WT censoring > 0.90 (47.1% at > 0.80; Table 3), versus 21.2% for cytotoxic agents — a 1.76× enrichment. The enrichment was stronger in a second GDSC panel: 29.4% targeted vs 8.3% cytotoxic (3.5×), indicating the geometry is not GDSC2-specific.

Table 3. Census of censoring geometry (non-informative fraction = WT censoring > 0.90; insufficient = a cell with < 5 lines). The non-informative fractions here use the non-silent mutation definition; the hotspot-defined value reported in the main text is 36.4%.

"Insufficient" here is among all drug × gene × lineage units; among the 200 hotspot-mutation candidates, 45.0% are insufficient (see the panoramic survey).

Panel	non-informative fraction (targeted)	non-informative fraction cytotoxic	insuff. among all units (targeted)	insuff. among all units (cytotoxic)
GDSC2	37.4%	21.2%	94.4%	82.0%
GDSC1	29.4%	8.3%	95.0%	82.8%

Scope note. This census operates on single-lineage mutant-vs-WT pairs (two cells), not the four-cell mutation×lineage interaction; the fraction of non-informative four-cell interactions is not directly measured here.

Table 4. Sensitivity of the bottleneck to the minimum-cell threshold (c* = 0.90).

min_n	insuff. (targeted)	non-informative fraction (targeted)	insuff. (cytotoxic)	non-informative fraction (cytotoxic)
3	88.8%	40.0%	69.9%	21.9%
5	94.4%	37.4%	82.0%	21.2%
10	98.4%	40.3%	92.4%	20.7%
15	99.3%	41.0%	96.7%	18.9%

Hotspot refinement. Using CCLE hotspot annotations (2.9% of non-silent calls) instead of all non-silent mutations, non-hotspot variants were *excluded* (assigned to neither cell) rather than reassigned to WT; the censoring imbalance $r_{WT} - r_{mut}$ became positive (median 0.083, 62.7% of units with mutant more sensitive than WT), consistent with the near-zero imbalance under the non-silent definition being an artifact of mixing non-hotspot (insensitive) mutants into the mutant cell.

Finally, two supplementary analyses refine the estimator comparison and the identifiability conclusion. As the WT-cell contrast δ_W varies (δ_M = 1.8 fixed), the mutant-only rank test's power stays flat while the OLS interaction power falls, confirming that the estimators target different estimands (Figure 4). The sharp identified interval for the dabrafenib interaction contains 0 only for an extrapolation bound B ≥ ln 6.6 (Figure 5).

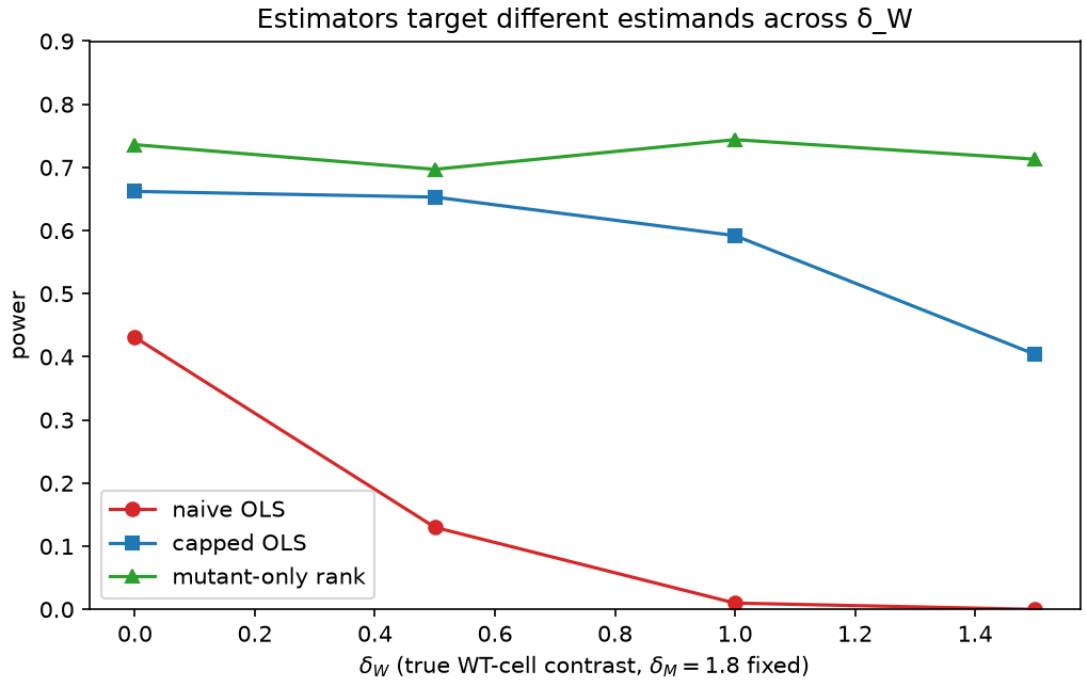


Figure 4

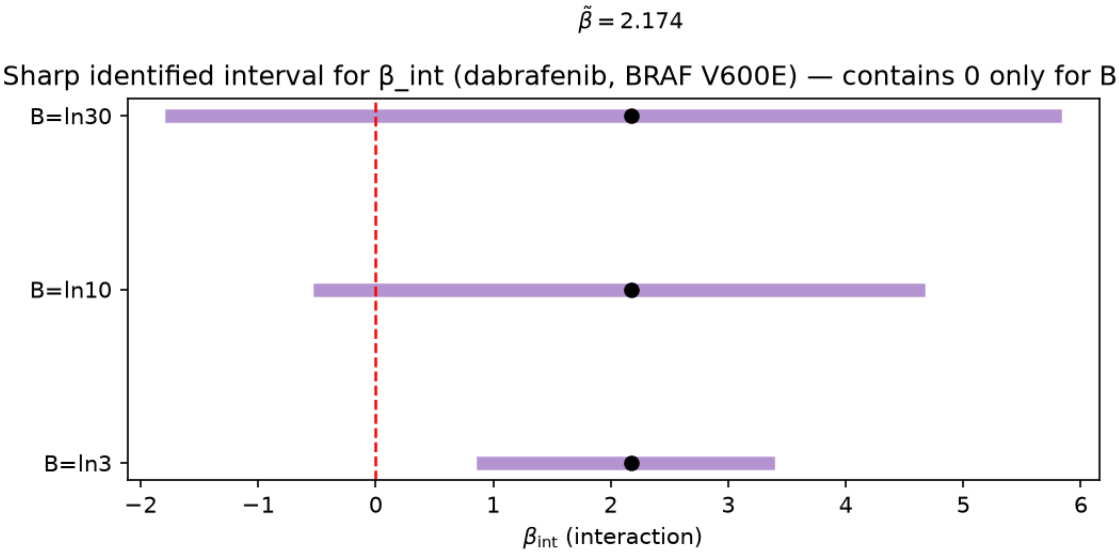


Figure 5

Construct validity: marginal extrapolation readout and effect-size divergence

Scope note. The positive controls in Methods are within-lineage (mutant vs WT) *marginal* contrasts, not 2×2 interactions. This section therefore demonstrates how extrapolation distorts the *marginal* readout under differential censoring; it does not bear construct validity for the interaction estimand itself.

Recall. Across the seven point-mutation positive controls, the naive estimator recalled 6/7 (core 4/4), capped 5/7, and rank 5/7 (Table 5). The single boundary case (nutlin-3a/TP53, capped β = 0.45 just below the floor) accounts for the naive vs capped difference; at n=7 this is not a statistical difference and we do not claim the naive estimator is "more sensitive".

Table 5. Positive-control recall (✓ = correct direction + one-sided p < 0.05 + effect-size floor).

Control	n_mut	n_WT	naive β	capped β	MWW δ
BRAF V600E / Dabrafenib / skin	38	15	−5.04 ✓	−3.06 ✓	−0.89 ✓
BRAF V600E / Trametinib / skin	40	15	−1.16 ✓	−1.11 ✓	−0.32 ✓
TP53 / Nutlin-3a / pan	659	307	+2.19 ✓	+0.45 ✗	+0.66 ✓
EGFR L858R / Afatinib / lung	5	172	−2.10 ✓	−2.13 ✓	−0.41 ✗
NRAS Q61 / Trametinib / skin	10	47	✗	✗	✗
PIK3CA / Alpelisib / breast	13	34	−2.08 ✓	−0.94 ✓	−0.75 ✓
KRAS G12/13 / Trametinib / lung	28	156	−1.90 ✓	−0.85 ✓	−0.53 ✓

Recall: naive 6/7, capped 5/7, MWW 5/7. (MWW δ is the Hodges–Lehmann estimate of the location shift.) (n is under the positive-control lineage filter; the four-cell n in the case dissection and the Appendix differ by lineage-filter details.)

Effect-size inflation. The more informative evidence is not recall but effect magnitude: the naive estimator's coefficient was 1.6–4.9× larger than the capped coefficient in 4 of 6 comparable controls (near 1× in the remaining two) (BRAF V600E: −5.04 vs −3.06; nutlin-3a: +2.19 vs +0.45; PIK3CA: −2.08 vs −0.94; KRAS: −1.90 vs −0.85). The naive estimator's extrapolation of resistant WT lines beyond the tested range produces coefficients larger than the capped estimator in the amplified subset (1.6–4.9×), so its higher recall accompanies larger coefficients, consistent with an 'extrapolation amplifies' mechanism

— an interpretation, not a causal claim established by the data; no ground truth is available to confirm which is closer to the true effect, so the divergence is a relative, not absolute, claim. This is the empirical realisation of the mechanistic claim: *extrapolation reads the censoring structure as signal — the bias direction is set by the censoring structure, not by the estimator's intent*.

Operational rules (R0–R5)

- **R0 (two-level gate, run first):** check per-cell n (insufficient cells do not enter inference), *then* check WT censoring ($> c^* \Rightarrow$ interaction non-informative, report the conditional-lineage contrast instead). Reversing the order leaves 94% of cells never reaching the censoring step.
- **R1a:** extrapolated IC50 point estimates are unreliable under this geometry; report them only alongside the censoring fraction.
- **R1b:** capped screening values are usable, but report the censoring fraction and the between-group Δcensoring.
- **R2:** on capped data, prefer rank/permutation tests; model-based standard errors require a heteroscedasticity-robust (sandwich) correction and are not recommended at small n.
- **R3:** prefer the censoring-probability interaction (E6) or the conditional-lineage rank contrast (mutant-only Mann–Whitney).
- **R4:** report the censoring matrix (per-cell n + censoring rate) as mandatory.
- **R5:** attach exact power to binary/tail endpoints.

Boundary note. Capping yields a per-cell *mean* lower bound, but the interaction is a signed ($\pm$) combination of cell means and has no such ordering; the naive and capped interaction estimates therefore bound neither the true interaction nor its sign. (nutlin-3a illustrates the capped estimator compressing the effect below the floor.)

Discussion

Bias direction is set by the censoring structure. The apparent tension between two failures of the naive estimator — inflated $\hat{\sigma}$ (false negatives) and inflated β (false positives) — resolves once extrapolation is seen as a single operator that scales *both* the numerator and the denominator of the test statistic. Under **differential** censoring (the WT cell is resistant and extrapolated far beyond the tested range), the extrapolation amplifies the effect in the same direction as the true signal, inflating β (positive controls, see Results). Under **non-differential** censoring (censoring unrelated to the mutation), it adds symmetric noise, inflating $\hat{\sigma}$ and deflating power (negative controls, see Results, FPR ≈ 5%). The claim — *extrapolation reads the censoring structure as signal; the bias direction is set by the censoring structure, not by the estimator's intent* — is therefore supported from both the simulation and the real-data sides, and it is exactly the mechanism behind Proposition 1's sign-reversal condition.

Why the mutant-only rank test wins. It discards the censored WT cells entirely, uses only the order information (which survives capping), and estimates a *different, identifiable* quantity — the conditional-lineage contrast — rather than the non-informative interaction. It is invariant to *WT-cell* censoring (which it discards) but not to mutant-cell censoring, which creates ties and loses order information, and its power still depends on the mutant-cell sample size. This is why it was optimal for its target (δ_M) in every cell of the two-dimensional grid (FPR 0.045, power 0.728) and why it is the natural primary endpoint once the interaction is acknowledged as non-informative.

The three-level bottleneck reframes the problem. The interaction failure is usually attributed to censoring alone. Our survey shows it is preceded by two harder filters: 94.5% of targeted-agent units carry no hotspot mutation in the queried lineage, and of the remaining candidates 45% are sample-size-insufficient (robust across thresholds 88.8–99.3%, reproduced in GDSC1). The censoring bottleneck (36% targeted vs 21% cytotoxic in GDSC2, 29% vs 8% in GDSC1) is the third. The two-level gate (R0) reflects this ordering: check n first, then censoring.

Relation to prior work. The censoring of IC50 and the robustness of AUC are established [3,7,5,6]. Tobit/censored regression is a standard tool but, as we show, fails under >90% single-cell censoring (17–19% non-convergence, coefficient bias up to $+0.86$, whose sign and magnitude depend on the effective WT cell size). This non-convergence is a finite-sample boundary event: the probability that a WT cell is fully censored is $\pi^n = 0.945^{17} = 0.38$ (skin-WT) and $0.945^{40} = 0.10$ (CRC-WT), and the estimator behaves once the effective cell size $m_l = n_l(1 - \pi_l) \gtrsim 5$ — the same admission condition as Corollary 1, which this observation therefore reinforces rather than contradicts. The distinctive contribution is the shift of the unit of analysis from the marginal drug–gene association to the 2×2 interaction, and the demonstration that its identifiability is governed by *per-cell* (not aggregate) censoring. Independent replication in a third panel, such as the PRISM viability screen [15], would further test the generality of this geometry.

Scope. The claims are restricted to selectively targeted agents in strong-phenotype settings (the T1 statement holds for these); we do not claim the failure for cytotoxic agents, whose WT cells are not near-totally censored. The survey is descriptive; the BRAF V600E case is post-hoc and exploratory.

Limitations

1. **Lineage–background confound.** The lineage contrast is a composite of tissue identity and its correlated molecular background (e.g., mutation co-occurrence); we cannot decompose these, and the mutation-stratified analysis is a partial, not a complete, correction.
2. **Small-cell shape estimation.** With $n = 6$ mutant lines in the BRAF case, the distributional shape within the mutant cell is not estimable; the rank test is preferred precisely because it makes no shape assumption.
3. **bounded-IC50 transform is rank-equivalent to IC50.** The transform is a deterministic monotone function of IC50 (fixed Hill slope), so it carries no order

information beyond IC50; its apparent robustness comes from boundedness and variance stabilization rather than from separating potency and efficacy.

4. **Mutation definition.** The survey uses non-silent mutations for the mutant cell; the hotspot refinement shows the censoring imbalance is sensitive to this definition (median $r_{WT} - r_{mut}$ from 0.000 to 0.083). The primary non-informative fraction is unaffected because it conditions only on the WT cell, whose definition is unambiguous.
5. **Positive-control set is small.** With $n = 7$ positive controls, the recall differences (6/7 vs 5/7) are not statistically distinguishable; the effect-size inflation, not the recall count, carries the claim.
6. **Post-hoc case.** The BRAF V600E dissection is exploratory; the confirmatory layer is the panoramic identifiability survey.

Conclusions

Pharmacogenomic interaction tests are routinely reported on a 2×2 mutation × lineage contrast whose two WT cells are near-totally censored for selective agents. Under this geometry the interaction is practically non-informative, and the two standard workarounds distort inference in opposite directions that are predictable from the censoring structure. We provide (i) the identifiability boundary and sign-reversal condition, (ii) a minimum-detectable-effect and cell-level admission heuristic, and (iii) an operational two-level gate: check per-cell sample size, then WT censoring; on censored data, use rank/permutation tests and report the censoring matrix; prefer the conditional-lineage rank contrast. These criteria constitute a methodological contribution rather than a critique: they give the analyst a checklist, implemented as an installable Python tool (censoring-check), for deciding before analysis whether a given mutation × lineage interaction is estimable. The three-level bottleneck — 94.5% without hotspot mutation, 45% sample-size-insufficient, 36.4% non-informative by censoring — is reproduced across panels and should be reported alongside any pharmacogenomic interaction claim.

Appendix

Proofs (Propositions 1–2, Corollary 1)

Proposition 1 (Partial identification under near-total WT censoring)

Assumptions. (A1) *Censoring.* Type-I right-censoring at c, observing $(y_{il} \wedge c, d_{il} = \mathbf{1}\{y_{il}^* > c\})$, $\pi_l = P(y^* > c)$. (A2) *Non-parametric model (for Prop 1(i)).* The four cell distributions F_l are unknown with a common latent support upper bound $\bar{y} = c + B$ ($B \in [0, \infty]$) and *no cross-cell restriction*. (A3) *Parametric model (for Prop 1(ii)-(iii)).* $y_{il}^* = \mu_l + \sigma\varepsilon_{il}$ with $\varepsilon_{il} \overset{\text{iid}}{\sim} \Phi$ (common location-scale normal). Write $s = (+1, -1, -1, +1)$, $\beta_{\text{int}} = \sum_l s_l \mu_l$.

Statement. - (i) *(non-parametric sharp set)* The sharp identified set for μ_l is $[\theta_l, \theta_l + \pi_l B]$, $\theta_l = E[y \wedge c]$; hence β_{int} has a sharp interval of width $W = B \sum_l \pi_l$ that does not shrink with n. As $\pi_{AW}, \pi_{BW} \to 1$, $W \to B(2 + \pi_{AM} + \pi_{BM})$: no consistent estimator exists, and any point estimate is determined by c rather than by the WT data. - (ii) *(parametric degeneracy)* With $F = \Phi$, μ_l is point-identified $\Leftrightarrow \pi_l < 1$; if a cell is fully censored the MLE diverges and the $1 - \alpha$ inversion set is a half-line, so β_{int} is bounded on one side only. - (iii) *(pseudo-limit of the capped OLS)* Regressing $y \wedge c$ on the saturated design, $\hat{\beta}_{\text{int}} \xrightarrow{p} \tilde{\beta}_{\text{int}} = \sum_l s_l m_l$, with $m_l = c - \sigma\Lambda(\zeta_l)$, $\zeta_l = (c - \mu_l)/\sigma$, $\Lambda(z) = \phi(z) + z\Phi(z)$; the c terms cancel and $\Lambda' = \Phi = 1 - \pi$.

Proof sketch. (1) Censoring identifies F on $(-\infty, c)$ and the mass π_l only; placing the censored mass at c versus $\bar{y}$ gives the Manski sharp bounds [11]. (2) β_{int} is a ± 1 linear functional over four independent cells with no cross-restrictions, so its sharp set is the Minkowski sum of the cell intervals. (3) $E[y \wedge c] = \mu\Phi(\zeta) - \sigma\phi(\zeta) + c(1 - \Phi(\zeta)) = c - \sigma\Lambda(\zeta)$, and the LLN under the saturated design gives (iii). ■

Proposition 2 (Fisher information and MDE lower bound)

Statement. Under (A1), (A3) with $F = \Phi$ and σ known, the per-cell Fisher information for μ_l is $I_1(\pi_l) = h(\pi_l)/\sigma^2$, where

$$h(\pi) = \Phi(\zeta) - \zeta\phi(\zeta) + \frac{\phi(\zeta)^2}{1 - \Phi(\zeta)}, \quad \zeta = \Phi^{-1}(1 - \pi),$$

strictly decreasing, $h(0) = 1$, $h(0.2) = 0.96$, $h(0.90) = 0.36$, $h(0.95) = 0.23$. The four-cell information matrix is diagonal, so for any (locally asymptotically) unbiased estimator

$$\text{Var}(\hat{\beta}_{\text{int}}) \geq \sigma^2 \sum_l \frac{1}{n_l h(\pi_l)} \equiv \frac{\sigma^2}{n_{\text{eff}}},$$

and the minimum detectable interaction effect (two-sided α, power $1 - \gamma$) is

$$\text{MDE} = \kappa\sigma \left(\sum_l [n_l h(\pi_l)]^{-1} \right)^{1/2} = \kappa\sigma/\sqrt{n_{\text{eff}}}, \quad \kappa = z_{1-\alpha/2} + z_{1-\gamma} \approx 2.80.$$

A design-robust *heuristic* (no tail extrapolation) replaces $h(\pi_l)$ by $1 - \pi_l$; because $h(\pi) > 1 - \pi$ for $\pi > 0$, this underestimates the information and is conservative for the MDE, so $n_{\text{eff}}^{\text{rob}} = (\sum_l m_l^{-1})^{-1}$ with $m_l = n_l(1 - \pi_l)$ dominated by the smallest m_l.

Proof sketch. (1) $\partial_\mu \ell = (1 - d)\frac{y - \mu}{\sigma^2} + d\frac{\phi(\zeta)}{\sigma(1 - \Phi(\zeta))}$; (2) the observed information contribution sums a truncated-variance term and a censoring-indicator term to give h; (3) under the saturated cell-means parameterization $\mathcal{I}$ is diagonal and $\beta_{\text{int}} = s^\top \mu$, giving the CRLB and the MDE by the normal power function. ■

Corollary 1 (cell-level admission heuristic)

Given (n_l, π_l, c, σ) and a target resolution Δ^* (log units, e.g. $\Delta^* = 1$): 1. **Admission (design-robust):** as a planning heuristic, interaction inference is practical only when $\sum_l m_l^{-1} \leq (\Delta^*/\kappa\sigma)^2$. A sufficient per-cell threshold is $m_l = n_l(1-\pi_l) \geq 4(\kappa\sigma/\Delta^*)^2$ for all l ($\sigma = 1$, $\kappa = 2.80$, $\Delta^* = 1 \Rightarrow m_l \geq 32$; $\Delta^* = 2 \Rightarrow m_l \geq 8$). 2. **Inverting for WT sample size:** $\pi_W = 0.93 \Rightarrow n_W \geq 32/0.07 \approx 460$; panels have $n_W \leq 50 \Rightarrow m_W \leq 3.5 \Rightarrow \mathrm{MDE} \geq 1.5$ log units — such a 2×2 cannot support interaction inference. 3. **Reporting rule:** when admission fails, report the sharp interval (width $B\sum_l \pi_l$) with π_l and m_l, or a censoring-aware quantity; prefer a censoring-aware quantity over naive/capped OLS point estimates or p-values, which are unreliable under this geometry.

Sign-reversal condition (capped OLS pseudo-limit)

Write $\delta_M = \mu_{AM} - \mu_{BM}$, $\delta_W = \mu_{AW} - \mu_{BW}$, $\beta_{\text{int}} = \delta_M - \delta_W$. By Prop 1(iii) and the mean-value theorem,

$$\tilde{\beta}_{\text{int}} = \bar{\Phi}_M \delta_M - \bar{\Phi}_W \delta_W, \quad \bar{\Phi}_g = \Phi(\tilde{\zeta}_g) \in [\min_{l \in g}(1-\pi_l), \max_{l \in g}(1-\pi_l)].$$

A sufficient (and, given $\bar{\Phi}$, necessary) condition for reversal is $\delta_M \delta_W > 0$ and $\bar{\Phi}_W/\bar{\Phi}_M < |\delta_M/\delta_W| < 1$, under which $\mathrm{sign}(\tilde{\beta}_{\text{int}}) = -\mathrm{sign}(\beta_{\text{int}})$. A conservative form using observed censoring rates is $(1-\pi_{W,\min})/(1-\pi_{M,\max}) < |\delta_M/\delta_W| < 1$. ■

Numerical demonstration of sign reversal

Homoscedastic DGP ($\sigma = 1$) with cells $n = (40,40,40,40)$, mutant cells at 20%/15% censoring and WT cells at 90%/78% censoring, $\delta_M = 0.2 < \delta_W = 0.5$ (true interaction -0.30; $\bar{\Phi}_W/\bar{\Phi}_M \approx 0.15 < 0.4$): the capped OLS converged to $+0.088$ (66.8% sign reversal), while the unbiased extrapolation stayed at -0.291. Under a common variance, a nonzero δ_W forces the two WT censoring rates to differ (here 90% vs 78%), and the capped OLS pseudo-limit reverses sign as predicted by the Corollary.

Rank-test failure DGP

With $\delta_M = 0$ and $\delta_W = 1$ (true interaction -1, driven entirely by the WT cells), the mutant-only rank test had power 0.050 (≈ α), the capped OLS interaction 0.157, and the un-censored OLS 0.467. The mutant-only rank test therefore targets δ_M only and fails when the interaction is driven by the WT cell — the stated boundary of rule R3.

Estimator targets differ across δ_W

With $\delta_M = 1.8$ fixed, the true interaction $\delta_M - \delta_W$ decreases as δ_W grows. The mutant-only rank test's power is flat (0.71–0.74) because it targets δ_M alone, while the OLS interaction power falls with the true interaction (naive 0.43 → 0.00, capped 0.66 → 0.40). The estimators target different estimands and their powers are not directly comparable.

Censoring-probability interaction (E6)

With $\delta_M = 0$ and $\delta_W = 1$ (interaction driven by the WT cell), the mutant-only rank test had power 0.046 (≈ α, blind to the WT cell), while the censoring-probability interaction (risk difference on the censoring indicator) had power 0.368. E6 retains the 2×2 structure, needs no extrapolation, and is the natural readout under the "WT fully resistant" geometry; it is complementary to the mutant-only rank test rather than a replacement.

Sharp interval on real data

For dabrafenib (BRAF V600E, CRC vs skin), the four cells are $n = (6,41,37,18)$ with censoring 17%/93%/22%/89% (strict V600E; the case dissection's $n = (6,40,38,17)$ differs by lineage-filter details and does not affect the conclusion). The capped OLS pseudo-limit is $\tilde{\beta} = 2.174$ and the sharp identified interval is $[-0.474,4.615]$ (width $B\sum_l \pi_l = 5.089$, $B = \ln 10$), which contains 0. The interval is B-dependent: at $B = \ln 3$ it is $[0.911,3.339]$, which excludes 0 (the interaction sign is identifiable as positive); it first contains 0 at $B \approx \ln 6.6$. The 'non-informative' conclusion therefore holds only for an extrapolation bound of $\gtrsim 6.6 \times$ MAX_CONC; for tighter bounds the interaction is sign-identified but not point-identified. A data-driven choice of B: across GDSC2, the naive extrapolation of censored observations beyond MAX_CONC has median $\ln 9.2 \approx 2.22$ and 90th percentile $\ln 41 \approx 3.72$; the default $B = \ln 10 \approx 2.30$ therefore sits at the median of GDSC's own extrapolation practice. Under this median-level B the interval contains 0 (non-informative); only under a tighter-than-median bound ($B \leq \ln 6.6$) does it exclude 0 (sign identified). The non-informative conclusion is thus supported at, and beyond, the extrapolation level GDSC itself routinely applies.

Tobit finite-sample boundary event

Varying the effective WT cell size $m_W = n_W(1 - \pi_W)$ with WT censoring fixed at 94.5%, the Tobit convergence rate rose from 88.6% ($m_W = 2$) to >98% ($m_W \geq 5$), and the interaction-coefficient bias shrank in magnitude from 0.183 ($m_W = 2$) to within ± 0.03 ($m_W \geq 10$). The non-convergence and finite-sample bias at small m_l therefore reflect $m_l = n_l(1 - \pi_l)$ too small, not a structural failure — consistent with Corollary 1's admission heuristic.

References

1. Garnett MJ, Edelman EJ, Heidorn SJ, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature. 2012;483:570–5.
2. Iorio F, Knijnenburg TA, Vis DJ, et al. A landscape of pharmacogenomic interactions in cancer. Cell. 2016;166:740–54.
3. Haibe-Kains B, El-Hachem N, Birkbak NJ, et al. Inconsistency in large pharmacogenomic studies. Nature. 2013;504:389–93.
4. Stransky N, Ghandi M, Kryukov GV, et al. Pharmacogenomic agreement between two cancer cell line data sets. Nature. 2015;528:84–7.

5. Vis DJ, Bombardelli L, Lightfoot H, et al. Multilevel models improve precision and speed of IC50 estimates. Pharmacogenomics. 2016;17:691–700.
6. Hafner M, Niepel M, Chung M, et al. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat Methods. 2016;13:521–7.
7. Safikhani Z, Smirnov P, Freeman M, et al. Revisiting inconsistency in large pharmacogenomic studies. F1000Res. 2016;5:2333.
8. Pozdeyev N, Yoo M, Mackie R, et al. Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies. Oncotarget. 2016;7:51619–25.
9. Jaeger S, Duran-Frigola M, Aloy P. Drug sensitivity in cancer cell lines is not tissue-specific. Mol Cancer. 2015;14:40.
10. Safikhani Z, El-Hachem N, Quevedo R, et al. Assessment of pharmacogenomic agreement. F1000Res. 2016;5:825.
11. Manski CF. Nonparametric bounds on treatment effects. Am Econ Rev. 1990;80:319–23.
12. Tobin J. Estimation of relationships for limited dependent variables. Econometrica. 1958;26:24–36.
13. Ghandi M, Huang FW, Jané-Valbuena J, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019;569:503–8.
14. Sharifi-Noghabi H, Jahangiri-Tazehkand S, Smirnov P, et al. Drug sensitivity prediction from cell line-based pharmacogenomics data: guidelines for developing machine learning models. Brief Bioinform. 2021;22:bbab294.
15. Corsello SM, Nagari RT, Spangler RD, et al. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. Nat Cancer. 2020;1:235–48.
16. Okyere D, Bravo-Merodio L, Xu Y, et al. Drug response in the era of precision medicine: a methodological review. Comput Struct Biotechnol J. 2025;27:5503–20.
17. Lenhof K, Eckhart L, Rolli L, et al. Reliable anti-cancer drug sensitivity prediction and prioritization. Sci Rep. 2024;14:62956.
18. Corcoran RB, Ebi H, Turke AB, et al. EGFR-mediated re-activation of MAPK signaling contributes to insensitivity of BRAF mutant colorectal cancers to RAF inhibition with vemurafenib. Cancer Discov. 2012;2:227–35.
19. Prahallad A, Sun C, Huang S, et al. Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. Nature. 2012;483:100–3.

Declarations

Ethics approval and consent to participate: Not applicable (this study used only publicly available cell-line data and did not involve human or animal subjects).

Consent for publication: Not applicable.

Availability of data and materials: All data analyzed in this study are publicly available. GDSC1/GDSC2 drug sensitivity data are from the Genomics of Drug Sensitivity in Cancer portal (https://cellmodelpassports.sanger.ac.uk/downloads). CCLE mutation calls (`CCLE_mutations.csv`) and `sample_info` are from DepMap release 26Q1 (`CCLE_mutations.csv`, `sample_info.csv`; https://depmap.org). All analysis scripts (Python) reproduce every figure and table; the code is available at https://github.com/kavanabj-ops/censoring-geometry2026.

Competing interests: The authors declare no competing interests.

Funding: This research received no external funding.

Authors' contributions: JB, TW and CS conceived the study; JB, TW and CS designed the methodology; JB wrote the software and performed the formal analysis; JB and TW curated the data and interpreted the results; JB drafted the manuscript; TW and CS reviewed and edited the manuscript; CS supervised and administered the project. All authors read and approved the final manuscript.

Acknowledgements: Not applicable.

Figure legends

Figure 1. Residual-SD ratio against cell counts at WT censoring 94.5%.

Figure 2. Censoring geometry of the dabrafenib BRAF V600E example.

Figure 3. Three-level bottleneck for targeted agents in GDSC2.

Figure 4. Power across δ_W ($\delta_M = 1.8$ fixed): the mutant-only rank test is flat (targets δ_M), while the OLS interaction power falls with the true interaction.

Figure 5. Sharp identified interval for β_{int} (dabrafenib, BRAF V600E): it contains 0 only for $B \geq \ln 6.6$.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- floatimage1.png