

Illustrative analysis. The sponsor, compound, study and site identifiers are fictional. Patient-level data were simulated, and the analyses were run on the simulated dataset.

SIGNAL RESILIENCE MAP · REPORT SR-0001 · V1.2

Does the X-score signal for AX-201 survive scrutiny?

Nordvik Therapeutics (fictional) plans to enrich its Phase II for patients with an X-score of 10% or more. We stress-tested that signal before the next trial is built around it.

Study	AX-201-101, Phase Ib/II expansion	Population	2L+ NSCLC after anti-PD-(L)1
N	86 treated · 65 response-evaluable with X-score	Endpoint	Objective response rate (RECIST 1.1)

SIGNAL AS PRESENTED TO THE BOARD 27% vs 7% ORR in X-score ≥10% (6/22) vs <10% (3/43). Difference +20 pp (95% CI +1 to +41). Unconfirmed responses, response-evaluable patients.	RESILIENCE VERDICT FRAGILE The difference falls to +7 pp when only confirmed responses count, and to +1 or +3 pp when the cut-off moves to 5% or 15%. Across 192 plausible specifications, only 11 had 95% CIs entirely above zero. The median difference is +6.5 pp. At the reported 10% cut-off, all 48 specifications had a positive ORR point estimate. Across all cut-offs, 141 of 192 did. The magnitude used to justify enrichment was not supported.
---	--

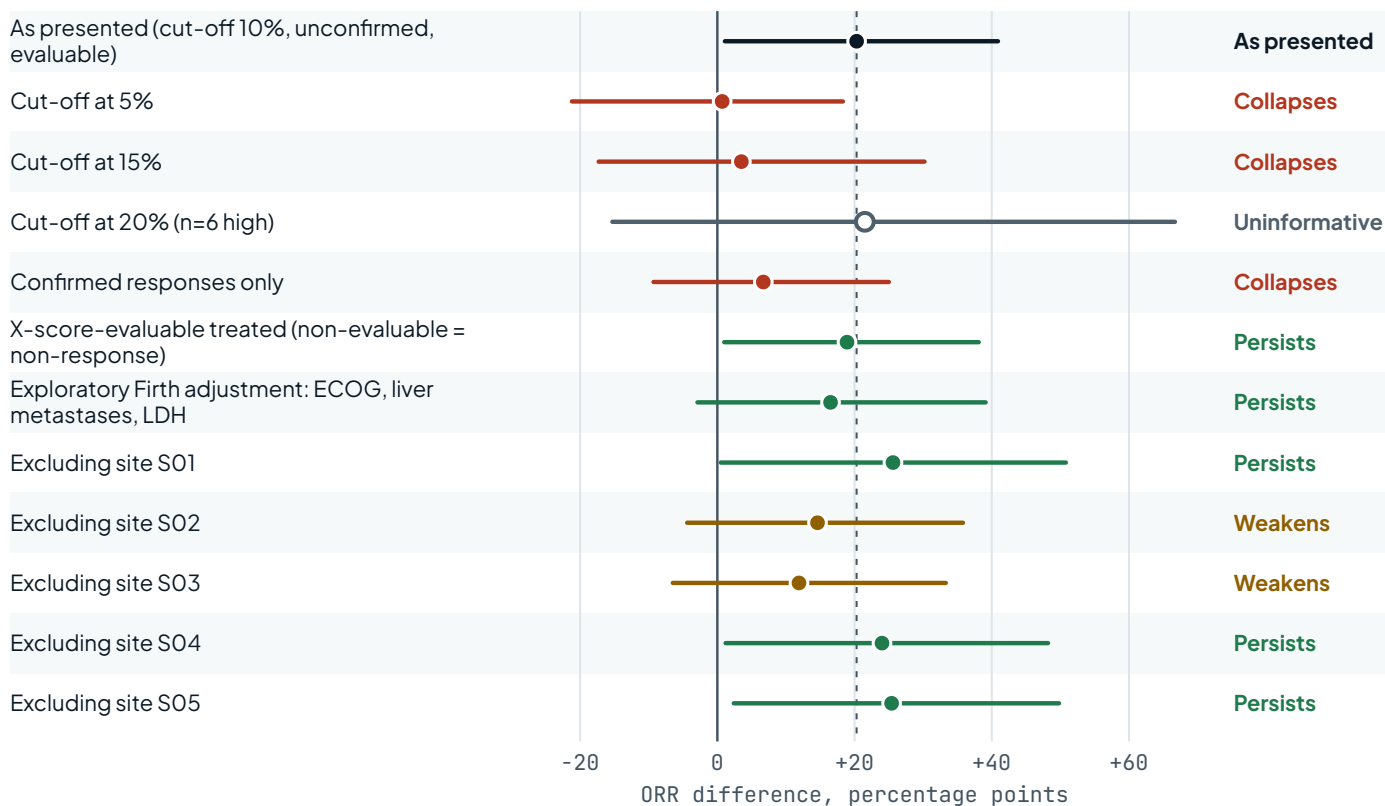
The map

Each assumption was changed one at a time from the analysis presented to the board. Rule fixed before the sensitivity-analysis results were inspected: **persists** if the difference stays at 80% or more of +20 pp, **weakens** at 50–80%, **collapses** below 50%, **uninformative** if a group has fewer than 10 patients. These are pre-specified descriptive resilience bands for communicating sensitivity to analytical choices; they are not statistical significance thresholds.

Persists X-score-evaluable treated population, non-evaluable response counted as non-response +19 pp (+1 to +38) Exploratory sparse-event adjustment (Firth) for ECOG, liver metastases and LDH +17 pp (–3 to +39) Excluding site S01, S04 or S05 Point estimates: +24 to +26 pp	Weakens Excluding site S02 +15 pp (–4 to +36) Excluding site S03 +12 pp (–6 to +33)	Collapses Confirmed responses only +7 pp (–9 to +25) Cut-off at 5% +1 pp (–21 to +18) Cut-off at 15% +3 pp (–17 to +30)	Uninformative Cut-off at 20% (6 patients met the ≥20% cut-off) +21 pp (–15 to +67)
--	--	---	--

One assumption at a time

Difference in ORR, X-score-high minus X-score-low, in percentage points with 95% bootstrap interval. The dashed line is the signal as presented.



Why the signal is fragile

The X-score-high ORR is based on six responses

Four of the six X-score-high responders have scores between 10 and 14, just above the cut-off. No patient scoring 15–19 responded.

X-SCORE	RESP.	ORR
0–4	2/15	13%
5–9	1/28	4%
10–14	4/10	40%
15–19	0/6	0%
≥20	2/6	33%

Half of the responses in the X-score-high group are unconfirmed

All three unconfirmed responses are in X-score-high patients, two of them at site S03. With confirmed responses only, ORR is 14% (3/22) vs 7% (3/43).

Baseline prognostic factors favoured the X-score-high group

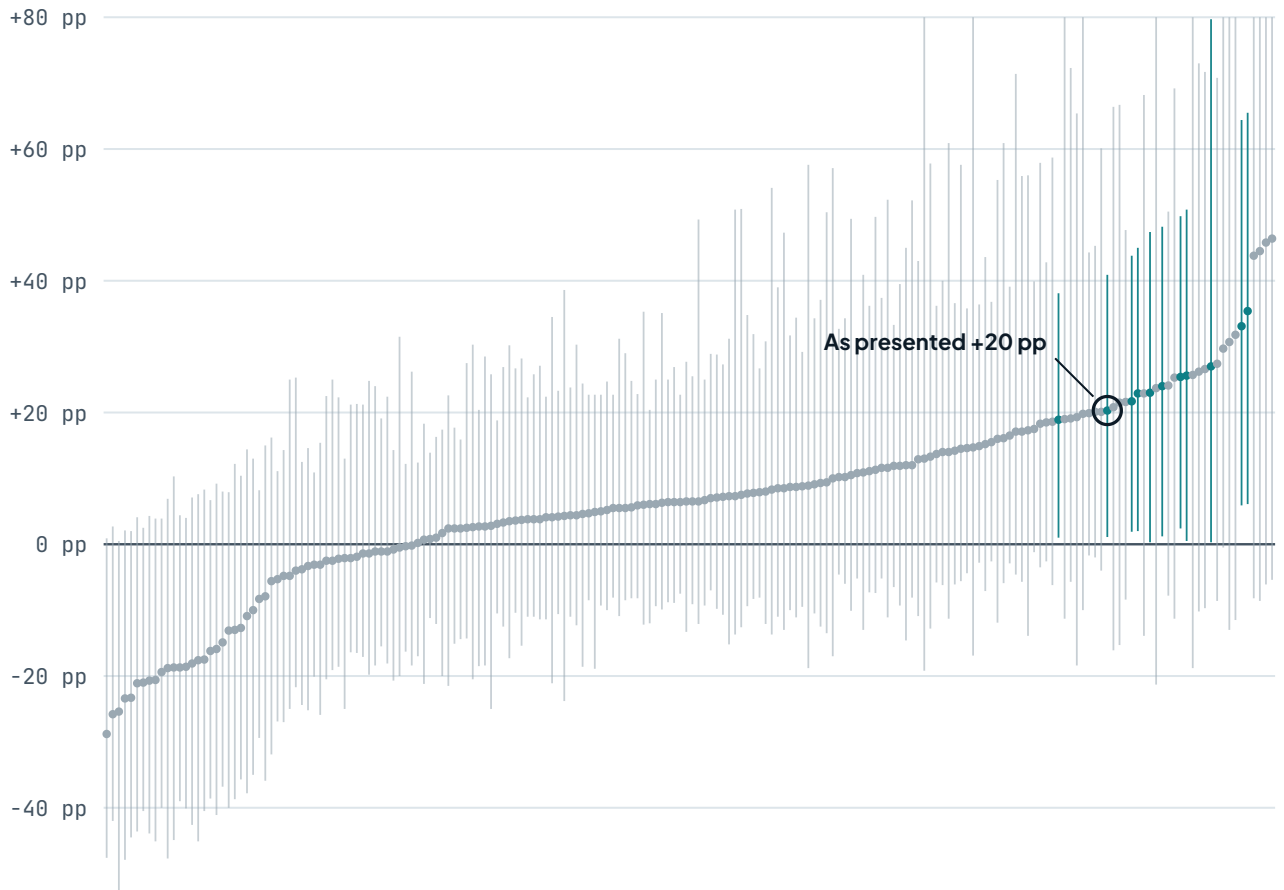
Liver metastases were present in 14% of X-score-high patients vs 42% of X-score-low patients. ECOG 1 was present in 27% of X-score-high patients vs 60% of X-score-low patients. Both are clinically relevant prognostic factors. Baseline prognostic imbalance may explain part of the observed difference.

As a continuous score, the odds ratio per 10 points is 1.9 (95% CI 0.7 to 5.8).

All 192 specifications, ranked

Every combination of cut-off (4), response definition (2), population (2), adjustment (none or exploratory Firth, 2) and site exclusion (6). Each dot is one specification with its interval. The grid below shows which choices produced it.

● Interval excludes zero (11) ● Interval includes zero (181) ○ Analysis presented to the board



What would change our confidence

Ranked by how directly each step addresses the main sources of decision uncertainty.

- 01 Confirm the three unconfirmed responses.** Obtain the confirmatory scans or a blinded independent central read. This alone determines whether the observed difference is approximately +20 pp or +7 pp.
- 02 Resolve the 13 missing X-scores.** Fifteen per cent of treated patients have no X-score. If archival tissue exists, scoring it could increase the number of patients with an X-score and help assess whether the missing data could bias the comparison.
- 03 Treat the 10% cut-off as a hypothesis.** It sits where four responders cluster. Pre-specify the continuous score as the primary biomarker analysis in Phase II and lock any cut-off before unblinding.
- 04 Do not run an enriched-only Phase II yet.** Under the illustrative one-sample design, 56 response-evaluable X-score-high patients are required to test an ORR of 22% against a 10% null rate (exact one-sided binomial test, α 0.05, 80% power; at least 10 responses required). Allowing for the 14% non-evaluable rate seen here, plan for about 66 treated X-score-high patients. Keep an X-score-low cohort and stratify by liver metastases and ECOG.

Analytical resilience is not a guarantee of clinical replication. This report shows how much the observed signal depends on analytical choices. It does not predict the Phase II result.

Methods and audit trail

Primary population: response-evaluable patients (at least one post-baseline scan) with an X-score, $n = 65$ of 86 treated. Alternative population: all treated patients with an X-score, $n = 73$; the 8 without a post-baseline assessment are counted as non-responders. The 13 treated patients without an X-score cannot be classified and are excluded from every comparison. Unconfirmed response: best overall response of CR or PR (RECIST 1.1); confirmed response: CR or PR confirmed at a subsequent assessment. Cut-offs tested: X-score ≥ 5 , ≥ 10 (as reported), ≥ 15 and ≥ 20 . Site specifications: all sites, or each of the five sites removed in turn. Adjusted specifications are an exploratory sparse-event adjustment: Firth-penalised logistic regression with X-score group, ECOG, liver metastases and LDH, followed by marginal standardisation over the analysed population to estimate an adjusted ORR difference; with 9 responses in the primary comparison these estimates are not confirmatory. Intervals are percentile bootstrap intervals based on 5,000 patient-level resamples per specification; resamples in which one X-score group is empty are discarded. The 192 specifications are the full grid of 4 cut-offs $\times$ 2 response definitions $\times$ 2 populations $\times$ 2 adjustment options $\times$ 6 site sets. The stress-test specification grid and descriptive resilience rules were fixed before the sensitivity-analysis results were inspected; the signal originally presented to the board was known beforehand. Code, data snapshot and outputs are versioned, so every number on this page can be regenerated.

Prepared by MedOncNORD AB · Signal Resilience · Illustrative analysis on simulated data
Sponsor, compound, study and site identifiers are fictional.