Endpoint-aware external evaluation of a published four-gene recurrence score in endometrioid endometrial carcinoma: a strict event-type analysis of TCGA-UCEC

Authors: [TO BE ADDED]

Affiliations: [TO BE ADDED]

Corresponding author: [NAME, ADDRESS, EMAIL — TO BE ADDED]

Abstract

Background

Published gene-expression scores are often described as externally validated when an expression matrix and a broadly related clinical field can be downloaded. In endometrial cancer, however, recurrence, progression, metastasis, and new primary malignancy are not interchangeable outcomes. We evaluated whether a published four-gene recurrence risk score could be transported to public endometrioid endometrial-cancer data while preserving the semantic boundary of the outcome.

Methods

We conducted a retrospective, endpoint-aware public-data evaluation of the fixed FN1/DUSP4/LEF1/SMAD9 score reported in GSE178671. GSE178671 was retained as the score-source reference cohort, and GSE216872 was audited for sample-level outcome availability. For TCGA-UCEC, we linked open BCR XML follow-up records to a UCSC Xena GDC STAR expression matrix and restricted the analysis to cases with explicit endometrioid histology and all four score genes. After a field-level audit, the author selected and froze an R2 strict endpoint: positive cases had a treatment-after-event status of YES and a case-level event type of Locoregional Recurrence or Distant Metastasis without any excluded event type; negative cases had a status of NO and no event type. New Primary Tumor, Locoregional Disease, Metastatic, untyped, ambiguous, and mixed-event records were excluded. Discrimination was summarized with ROC-AUC and average precision (PR-AUC), with 2,000 non-parametric bootstrap resamples.

Results

The GSE178671 matrix contained 49 evaluable samples (16 recurrence, 33 non-recurrence), and the fixed score showed apparent reference-cohort discrimination (ROC-AUC 0.930, 95% bootstrap CI 0.845–0.987; PR-AUC 0.852). GSE216872 provided 61 expression columns but no public sample-level recurrence mapping in the downloaded artifacts, so it could not yield a reproducible outcome analysis. The TCGA query returned 548 open clinical files and 585 expression columns; 545 cases matched all four genes, 409 had explicit endometrioid histology, and 329 cases met the R2 strict endpoint rules (18 positive, 311 negative). In this primary TCGA evaluation, ROC-AUC was 0.313 (95% CI 0.175–0.466) and PR-AUC was 0.041 against a 5.5% event prevalence. The pre-specified positive-type-priority sensitivity analysis included 331 cases (20 positive, 311 negative) and produced ROC-AUC 0.373 (95% CI 0.212–0.535) and PR-AUC 0.055. A legacy YES/NO composite event analysis yielded ROC-AUC 0.451 (95% CI 0.364–0.542) and was not interpreted as recurrence validation.

Conclusions

The published score was reproducible in its reference matrix but did not demonstrate transportable discrimination under the author-defined TCGA R2 strict endpoint. The principal methodological finding is that an accessible clinical event field cannot be silently relabeled as recurrence. Public-data external evaluation in gynecologic oncology should separate endpoint equivalence, histology, expression compatibility, and sample-level linkage before reporting model performance.

Keywords: endometrial cancer; endometrioid carcinoma; recurrence risk score; external evaluation; public data; TCGA; GEO; endpoint adjudication; transportability; reproducibility

Introduction

Endometrial carcinoma is biologically and clinically heterogeneous. Endometrioid tumors comprise a major histologic group, yet recurrence risk remains difficult to summarize across cohorts because studies differ in eligibility, treatment context, molecular measurement, follow-up definitions, and the way recurrence is recorded. Gene-expression signatures can provide a compact risk summary, but their apparent performance depends on whether the outcome measured in a validation cohort is genuinely equivalent to the outcome used during development.

The distinction is particularly important in public-data research. A downloadable expression matrix may be accompanied by a series-level statement that a study contains recurrent and non-recurrent patients without exposing a sample-to-label mapping. Similarly, an open cancer registry may contain a field for a new tumor event after initial treatment that combines progression, recurrence, metastasis, and new primary malignancy. Such fields are valuable clinical information, but they do not automatically constitute a pure binary recurrence endpoint. Treating them as interchangeable can produce an apparently complete external validation while estimating a different clinical quantity.

GSE178671 reported a four-gene score based on FN1, DUSP4, LEF1, and SMAD9 for recurrence-risk classification in endometrial cancer [1]. GSE216872 is a biologically relevant endometrioid recurrence study with publicly discoverable expression artifacts, but its downloadable files do not expose a reproducible sample-level recurrence label [2,3]. TCGA-UCEC offers broad molecular and clinical resources and is therefore an attractive candidate for external evaluation. Its legacy BCR XML files also contain a more granular new-neoplasm event-type field in addition to the broader treatment-after-event status [4,5].

We therefore designed an endpoint-aware evaluation with four safeguards. First, the original four-gene formula was frozen and never re-estimated. Second, endometrioid histology was required from an original field rather than inferred from the project name. Third, the TCGA endpoint was adjudicated from event status and event type, with mixed and ambiguous event records excluded from the primary analysis. Fourth, performance was reported with both ROC-AUC and PR-AUC because the strict endpoint was uncommon. Our objective was not to develop a new classifier or claim clinical utility, but to determine what external claim is supported by the public artifacts after outcome semantics have been audited.

Methods

Study design and protocol chronology

This was a retrospective methodological evaluation of public de-identified data. The score formula, target population, data-access boundaries, and analysis code were maintained in a versioned local study package. The project first performed a data-feasibility audit. The discovery that TCGA legacy XML contained an event-type field led to an author-level G2 choice of the R2 strict endpoint. The R2 rule was then frozen before the full performance analysis and manuscript reconstruction. Accordingly, the TCGA results are best interpreted as an endpoint-audited, field-informed external evaluation rather than a prospectively blinded validation study.

The study followed the reporting logic of STROBE and TRIPOD for prediction-model evaluation. The concepts, definitions, and boundary conditions were governed by a separate concept contract. G3 actions—repository upload, journal submission, external messaging, and publication—were outside the scope of this analysis.

Data sources and cohort roles

GSE178671 reference cohort

GSE178671 was downloaded from the NCBI Gene Expression Omnibus (GEO) series matrix. The cohort was used only to reproduce the published score behavior in the accessible matrix and was not counted as external validation. The matrix contained 49 samples with four score genes and sample titles that distinguished recurrence from non-recurrence. The matrix-level sample count, rather than the larger series-level description, was used for reproducibility.

GSE216872 label-availability audit

GSE216872 was assessed using the public GEO series matrix and a publicly downloadable supplementary expression file. The expression artifact contained 61 sample columns and the four score genes were retrievable. The downloaded artifacts did not contain a sample-level recurrence/non-recurrence variable linked to those expression columns. The published series-level event count was therefore not converted into labels without a public sample-to-label mapping.

TCGA-UCEC external evaluation cohort

Open TCGA-UCEC BCR XML clinical supplement files were obtained through the GDC open-access workflow. The query returned 548 clinical files linked to 548 cases. Four-gene expression values were obtained from the publicly downloadable UCSC Xena GDC STAR matrix, which was treated as a processed log2(x+1) expression source. Expression columns were linked to clinical records by TCGA submitter case identifiers. If a case had multiple expression samples, a primary-tumor sample with sample-type code 01 was preferred; if more than one remained,

the lexicographically first sample was retained. The number of multi-sample cases was recorded rather than silently discarded.

The primary TCGA population required an original histology field explicitly identifying endometrioid carcinoma. Non-endometrioid, mixed, and unknown histology records were not included in the primary endpoint analysis.

Frozen four-gene score

For every evaluable sample, the score was calculated as:

$$RRS = -21.14 + 1.02 \times FN1 + 1.07 \times DUSP4 + 0.6211 \times LEF1 + 0.8832 \times SMAD9.$$

The four genes were mapped using fixed Ensembl identifiers: FN1, ENSG00000115414; DUSP4, ENSG00000120875; LEF1, ENSG00000138795; and SMAD9, ENSG00000120693. The score was calculated only when all four values were finite and uniquely mapped. No cohort-specific centering, scaling, batch correction, coefficient recalibration, gene selection, or outcome-dependent transformation was performed. No validation-set cutpoint was optimized; the score remained continuous for discrimination analyses.

Endpoint adjudication

The TCGA endometrial follow-up form distinguishes whether a new tumor event occurred after initial treatment from the type of that event. The field vocabulary includes Locoregional Recurrence, Distant Metastasis, and New Primary Tumor [5]. The broader treatment-after-event status can contain clinically heterogeneous events and was therefore not treated as pure recurrence.

For the author-confirmed R2 strict endpoint, the XML records were aggregated at the case level. A case was positive only if the event status included YES, the union of

non-empty event types contained Locoregional Recurrence or Distant Metastasis, and the same case-level union contained none of New Primary Tumor, Locoregional Disease, or Metastatic. A case was negative only if the event status was NO and the event-type union was empty. Cases with missing or contradictory status, YES without a type, any excluded type, or an excluded mixed-type combination were omitted. This rule yielded the primary binary variable `r2_status`, with 1 indicating an R2 event and 0 indicating an R2 negative case.

Two sensitivity definitions were frozen before analysis. The positive-type-priority sensitivity analysis counted a case as positive when YES and at least one of the two positive types was present, even if another type was also present. The legacy composite analysis used the broader YES/NO treatment-after-event status and was explicitly labelled as a non-equivalent composite endpoint.

Statistical analysis

The primary prediction variable was the continuous RRS in its published direction. ROC-AUC was calculated from rank ordering. Because positive events were uncommon, average precision was also reported as PR-AUC together with the observed event prevalence. Confidence intervals were obtained from 2,000 non-parametric bootstrap resamples using random seed 20260906; resamples containing only one outcome class were excluded from interval calculation. We report sample size, event count, non-event count, prevalence, score medians, ROC-AUC, PR-AUC, and 95% bootstrap intervals.

No new model was trained. No multivariable adjustment, high-dimensional feature selection, verification-set threshold optimization, calibration model, decision-curve analysis, clinical net benefit analysis, or treatment interaction analysis was

performed. The number of primary events was too small for a stable new multivariable model and the R2 endpoint was not a time-to-event outcome.

Reproducibility and data governance

The analysis was implemented in Python 3.13.5 with NumPy and deterministic local parsers. The primary analysis script was `scripts/analyze_external_validation.py`; the R2 field audit was implemented in `scripts/audit_tcga_event_types.py`. Figures were generated by `scripts/build_r2_figures.R`. Input URLs, local file hashes, clinical-file counts, field mappings, exclusion reasons, and derived result tables were retained in the project package. The source manifest was frozen; the R2 field-semantics sources were recorded in a separate candidate addendum without changing the frozen base manifest.

Results

Public-data availability and cohort roles

The reference matrix and two candidate external resources produced different forms of evidence. GSE178671 supported a matrix-level reference reproduction. GSE216872 supported expression availability but not a reproducible sample-level endpoint. TCGA-UCEC supported both expression and clinical-field linkage after an event-type audit, but its primary endpoint was an author-defined recurrence-compatible event rather than pure recurrence.

Table 1. Cohort roles and endpoint eligibility.

Cohort	Expression artifact	Outcome artifact	Population/role	Decision
GSE178671	49 evaluable samples	16 recurrence, 33 non-	Published-score	Reference reproduction

Cohort	Expression artifact	Outcome artifact	Population/role	Decision
		recurrence	reference	only
GSE216872	61 sample columns	No public sample-level recurrence mapping	Endometrioid label-availability audit	No AUC generated
TCGA-UCEC	585 columns; 545 cases matched to four genes	Legacy event status plus event type	Explicit endometrioid external evaluation	R2 strict primary analysis

Reference-cohort reproduction

All 49 GSE178671 matrix samples were evaluable for the four-gene score and reference outcome. There were 16 recurrence and 33 non-recurrence samples. The median RRS was 1.600 among recurrence samples and −3.418 among non-recurrence samples. The apparent reference-cohort ROC-AUC was 0.930 (95% bootstrap CI 0.845–0.987), and PR-AUC was 0.852 (95% CI 0.645–0.980). These values describe the accessible score-source matrix and were not used as evidence of independent validation.

GSE216872 label audit

The GSE216872 supplementary expression file contained 61 sample columns with the four score genes. However, the public artifacts examined here did not contain a sample-level recurrence field linked to those columns. A series-level statement that the study included recurrent and non-recurrent patients cannot establish which expression column belongs to which outcome. Accordingly, GSE216872 was retained as a transparent label-availability audit and did not contribute an AUC or PR-AUC.

TCGA-UCEC sample flow and endpoint composition

The GDC query returned 548 clinical XML files. The expression matrix contained 585 columns, and 545 cases could be matched to all four genes. Twenty-seven cases had multiple expression columns; the pre-specified primary-tumor selection rule was applied. Of the 545 matched cases, 409 had explicit endometrioid histology, 114 were non-endometrioid, and 22 were mixed or otherwise not eligible for the primary population.

Among the 409 endometrioid cases, 311 had an explicit NO event status and no event type. Forty-five had missing/indeterminate event status with no type, two had YES without an event type, and the remainder contained event types that were excluded from the strict primary endpoint. The case-level event-type audit documented Locoregional Recurrence, Distant Metastasis, Metastatic, Locoregional Disease, New Primary Tumor, and three mixed-type combinations (Figure 4). After all strict rules were applied, 329 cases remained: 18 positive and 311 negative.

Table 2. TCGA-UCEC R2 strict endpoint adjudication.

Field-level category	Endometrioid cases
Matched to four genes	409
R2 strict negative: NO + no type	311
R2 strict positive: eligible event type	18
Missing/indeterminate status, no type	45
YES without event type	2
Excluded type or mixed event	33
R2 strict analysis set	329

The 33 excluded-type/mixed-event count represents case-level exclusion categories and is not intended to be interpreted as mutually exclusive raw XML rows. Multiple

follow-up records were collapsed to a case-level event-type union before adjudication.

Primary and sensitivity performance

In the R2 strict primary analysis, the score had ROC-AUC 0.313 (95% bootstrap CI 0.175–0.466) and PR-AUC 0.041 (95% CI 0.023–0.067) in 329 cases with an event prevalence of 5.5%. The median score was −12.272 in event-positive cases and −10.400 in event-negative cases. Thus, the observed score direction in TCGA was opposite to the direction seen in the GSE178671 reference matrix; the direction was not reversed for the primary analysis.

The positive-type-priority sensitivity analysis included 331 cases (20 positive, 311 negative). ROC-AUC was 0.373 (95% CI 0.212–0.535) and PR-AUC was 0.055 (95% CI 0.030–0.095), with 6.0% event prevalence. The legacy YES/NO composite analysis included 364 endometrioid cases (53 positive, 311 negative) and produced ROC-AUC 0.451 (95% CI 0.364–0.542) and PR-AUC 0.147 (95% CI 0.103–0.220). The composite estimate was reported for transparency and was not re-labelled as recurrence discrimination.

Table 3. Fixed-score discrimination results.

Analysis	Endpoint	n (positive/negative)	ROC-AUC (95% bootstrap CI)	PR-AUC (95% bootstrap CI)
GSE178671 reference	Recurrence/ non-recurrence	49 (16/33)	0.930 (0.845–0.987)	0.852 (0.645–0.980)
TCGA R2 strict	Recurrence-compatible event	329 (18/311)	0.313 (0.175–0.466)	0.041 (0.023–0.067)
TCGA positive-type priority	Recurrence-compatible event	331 (20/311)	0.373 (0.212–0.535)	0.055 (0.030–0.095)

Analysis	Endpoint	n (positive/negative)	ROC-AUC (95% bootstrap CI)	PR-AUC (95% bootstrap CI)
	sensitivity			
TCGA legacy composite	New-tumor event YES/NO	364 (53/311)	0.451 (0.364–0.542)	0.147 (0.103–0.220)

Figure 2 displays the ROC and precision–recall curves. Figure 3 shows the score distributions. The high apparent separation in the reference matrix was not reproduced under the TCGA R2 endpoint, and the low PR-AUC was close to the low event baseline expected in an imbalanced cohort.

Discussion

Principal findings

This endpoint-aware evaluation produced three findings. First, the published four-gene score was computable and showed strong apparent discrimination in the accessible GSE178671 reference matrix. Second, GSE216872 could not be used for a reproducible outcome analysis because the public expression artifacts did not provide a sample-to-recurrence mapping. Third, TCGA-UCEC contained sufficient open expression and clinical fields for a transparent event-type evaluation, but the score did not show transportable discrimination under the author-confirmed R2 strict endpoint.

The primary result should not be summarized as “the score failed recurrence validation” without qualification. The TCGA R2 endpoint is deliberately narrower than the broad treatment-after-event field, but it remains a field-derived recurrence-compatible event rather than a harmonized pure recurrence or survival endpoint. The statistically low ROC-AUC may reflect endpoint heterogeneity, platform and preprocessing differences, cohort differences, reverse score orientation, or a

genuine lack of transportability. The study was designed to expose these possibilities rather than collapse them into one causal explanation.

Endpoint equivalence is part of validation

External validation is not established by expression availability alone. It requires a defensible relationship between the development outcome and the validation outcome. The TCGA follow-up structure is clinically informative precisely because it distinguishes types of new tumor events. It also demonstrates why a broad YES/NO field cannot automatically be named recurrence. In the present data, the strict endpoint excluded cases with new primary tumor, locoregional disease, metastatic legacy values, untyped events, and mixed combinations. This reduced the analysis set but made the estimand explicit.

The same principle explains the treatment of GSE216872. A published event count is not a public sample-level label. Assigning seven recurrent cases to arbitrary expression columns would create a data-dependent outcome and invalidate the intended external evaluation. Keeping GSE216872 as a label-availability audit is therefore a positive reproducibility result, even though it produces no AUC.

Reference performance versus transportability

The reference matrix reproduced high apparent discrimination, whereas the TCGA R2 strict analysis showed an AUC below 0.5 in the published score direction. This contrast is scientifically important but not, by itself, proof of a biological reversal. The data sets differ in platform, population composition, clinical context, sampling and endpoint construction. Because the primary score direction was frozen, we did not flip the score to obtain a more favorable AUC. A direction-reversed analysis could be

considered a separate hypothesis in future work, but it would not rescue the prespecified external evaluation.

The R2 strict result also illustrates why PR-AUC is useful. With only 18 positive cases among 329, a model can have unstable precision–recall behavior even when ROC summaries appear interpretable. The observed PR-AUC of 0.041 was below the observed event prevalence of 0.055. We therefore avoid describing the score as clinically useful or as a validated patient-level risk tool.

Strengths

The study has several strengths. The score formula was fixed and traceable to the published source. The reference, unavailable, and external-evaluation cohorts were assigned distinct roles. Histology was read from original fields rather than inferred from project names. TCGA events were adjudicated at the case level with explicit exclusion rules. The primary and sensitivity endpoint definitions were preserved in the analysis output. Both ROC-AUC and PR-AUC were reported with deterministic bootstrap intervals. Finally, the local package retains raw acquisition records, field audits, scripts, figures and a submission-ready manuscript without making an external upload.

Limitations

First, the R2 endpoint was selected after a field-level feasibility audit, so the TCGA evaluation is field-informed and hypothesis-generating rather than a prospectively blinded validation. Second, the strict endpoint contained only 18 positive cases, limiting precision and precluding stable multivariable adjustment, calibration modeling and clinical utility analysis. Third, legacy XML records may contain multiple follow-up observations and historical values whose clinical context cannot be fully

reconstructed from the open artifacts; the case-level union rule is reproducible but not equivalent to adjudication by a treating clinician. Fourth, the R2 endpoint is not pure recurrence and does not establish DFS, PFS or OS performance. Fifth, the reference cohort was small and its apparent discrimination is not an unbiased out-of-sample estimate. Sixth, expression-scale compatibility was audited using available processed matrices but cannot remove all cross-platform differences. Seventh, the public GSE216872 artifacts may not include private or later-released clinical annotations. Finally, institutional data-governance and ethics language should be confirmed by the authors before journal submission.

Implications

For investigators using public gynecologic-oncology data, the practical sequence is outcome mapping, histology adjudication, expression compatibility, sample linkage, and only then model performance. When a public cohort lacks an equivalent sample-level endpoint, the appropriate conclusion may be that external validation is not estimable. When a related field is usable under a transparent operational rule, it should be reported as a distinct endpoint with its own limitations. This approach can turn a superficially simple signature-validation exercise into a more informative study of transportability and reproducibility.

Conclusions

The GSE178671 four-gene score was reproducible in its reference matrix but did not demonstrate transportable discrimination in the TCGA-UCEC R2 strict endpoint evaluation. The major contribution of this study is methodological: recurrence-score validation requires explicit endpoint equivalence, not merely a downloadable expression matrix and a clinically related field. The R2 strict result should be treated

as an endpoint-aware external evaluation with limited event counts, not as evidence of clinical utility or of pure recurrence prediction.

List of abbreviations

AUC, area under the receiver operating characteristic curve; GDC, Genomic Data Commons; GEO, Gene Expression Omnibus; PR-AUC, precision–recall area under the curve; RRS, recurrence risk score; TCGA, The Cancer Genome Atlas.

Declarations

Ethics approval and consent to participate

This study used publicly accessible, de-identified GEO and GDC/UCSC Xena artifacts and did not involve recruitment or direct contact with human participants. The authors should confirm the applicable institutional and journal requirements for secondary use of public data before submission.

Consent for publication

Not applicable.

Availability of data and materials

The input artifacts were obtained from NCBI GEO, the NCI Genomic Data Commons open clinical supplement, and the UCSC Xena GDC STAR expression hub. Accession roles, acquisition records, field mappings, hashes, analysis scripts and derived tables are included in the local study package. No public code or data repository upload has been performed in this workflow.

Competing interests

[TO BE CONFIRMED BY AUTHORS]

Funding

No external funding statement has been supplied. The authors should confirm the final funding declaration before submission.

Authors' contributions

The final author contribution statement requires author completion before submission.

Acknowledgements

[TO BE CONFIRMED BY AUTHORS]

References

1. de Andrade DA, da Silva LS, Laus AC, et al. A 4-gene signature associated with recurrence risk in patients with low and intermediate-risk endometrial cancer. *Front Oncol*. 2021;11:729219. doi:10.3389/fonc.2021.729219.

2. Gonzalez-Bosquet J, Devor EJ. Integration of genomic and clinical retrospective data to predict endometrioid endometrial cancer recurrence. *Int J Mol Sci*. 2022;23(24):16014. doi:10.3390/ijms232416014.

3. National Center for Biotechnology Information. GSE216872: Integration of genomic and clinical data to predict endometrioid endometrial cancer recurrence. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE216872. Accessed 6 Sep 2026.

4. National Cancer Institute Genomic Data Commons. Legacy Archive TCGA Tag Descriptions. https://gdc.cancer.gov/resources-tcga-users/legacy-archive-tcga-tag-descriptions. Accessed 6 Sep 2026.

5. The Cancer Genome Atlas. Endometrial Follow-Up Form. https://www.nationwidechildrens.org/-/media/nch/research/documents/tcga/endometrial-followup-form.ashx. Accessed 6 Sep 2026.

6. National Cancer Institute Genomic Data Commons. Data Submission Walkthrough: diagnosis and progression/recurrence fields. https://docs.gdc.cancer.gov/Data_Submission_Portal/Users_Guide/Data_Submission_Walkthrough/. Accessed 6 Sep 2026.

7. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162:55–63. doi:10.7326/M14-0697.

8. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement. *PLoS Med*. 2007;4:e296. doi:10.1371/journal.pmed.0040296.

Figure legends

Figure 1. Public-data cohort roles and R2 strict evaluation. The figure separates the GSE178671 reference reproduction, the GSE216872 label-availability audit, and the TCGA-UCEC R2 strict external-evaluation route. The TCGA route is shown as endpoint-aware evaluation rather than pure recurrence validation.

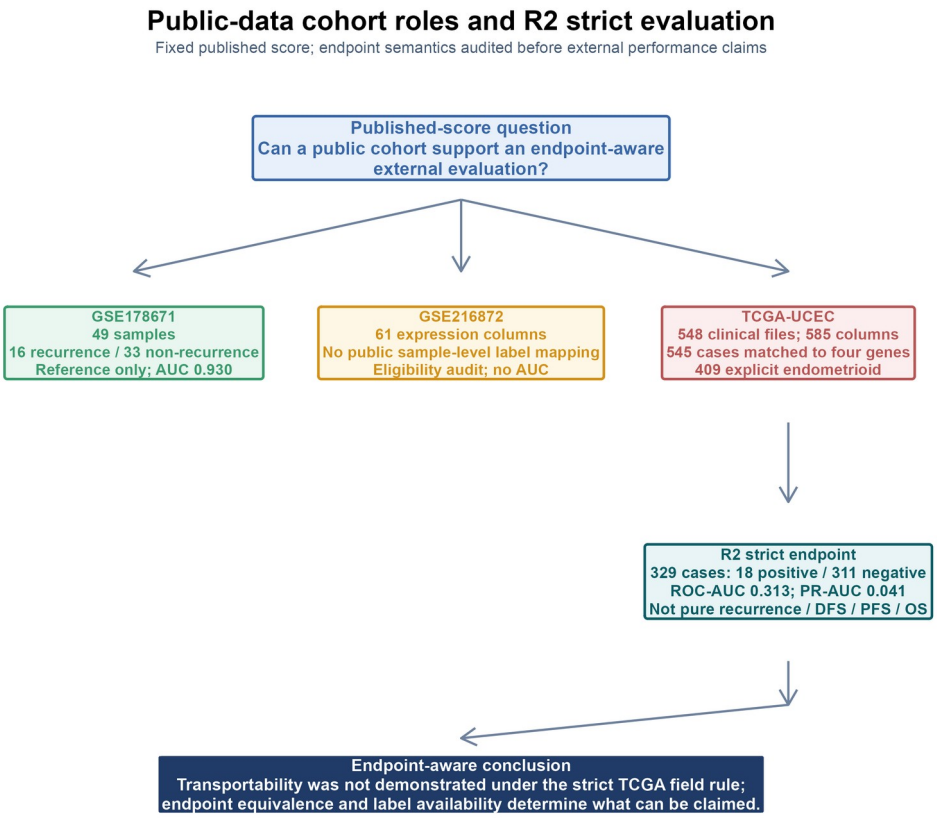


Figure 2. Fixed-score discrimination across the reference and TCGA endpoint definitions. ROC and precision–recall curves are shown for the GSE178671 reference cohort, TCGA R2 strict endpoint, positive-type-priority sensitivity endpoint, and the legacy composite event endpoint. The TCGA curves are not pooled with the reference curve.

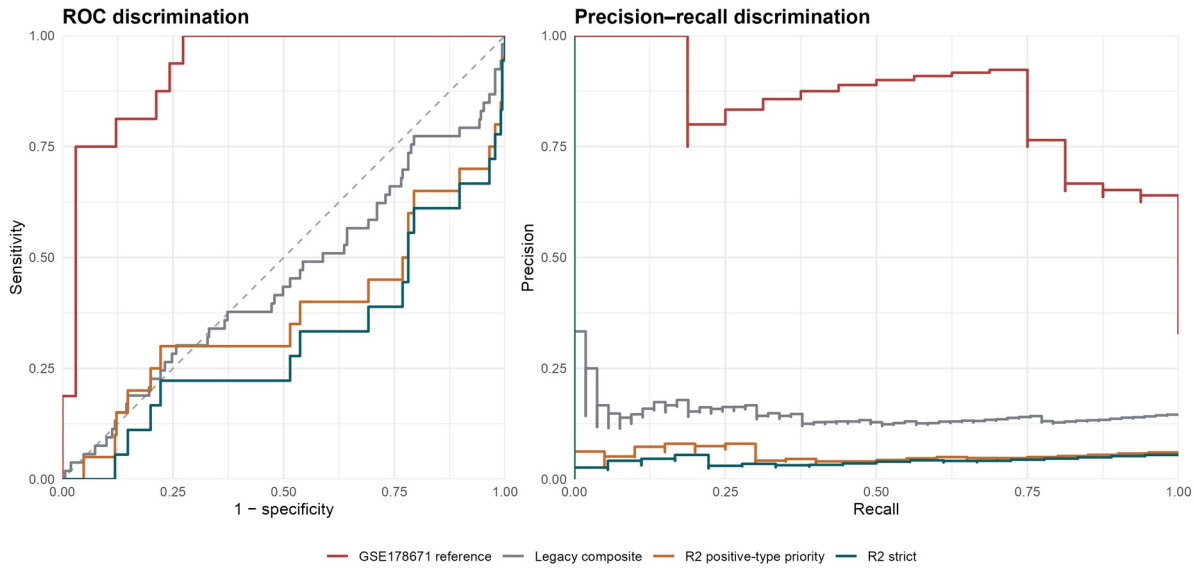


Figure 3. Frozen score distributions. Violin distributions, boxplots and individual observations are shown for the TCGA R2 strict event-positive/event-negative groups and the GSE178671 reference recurrence/non-recurrence groups. The vertical score scale is retained across groups; no post hoc reorientation was applied.

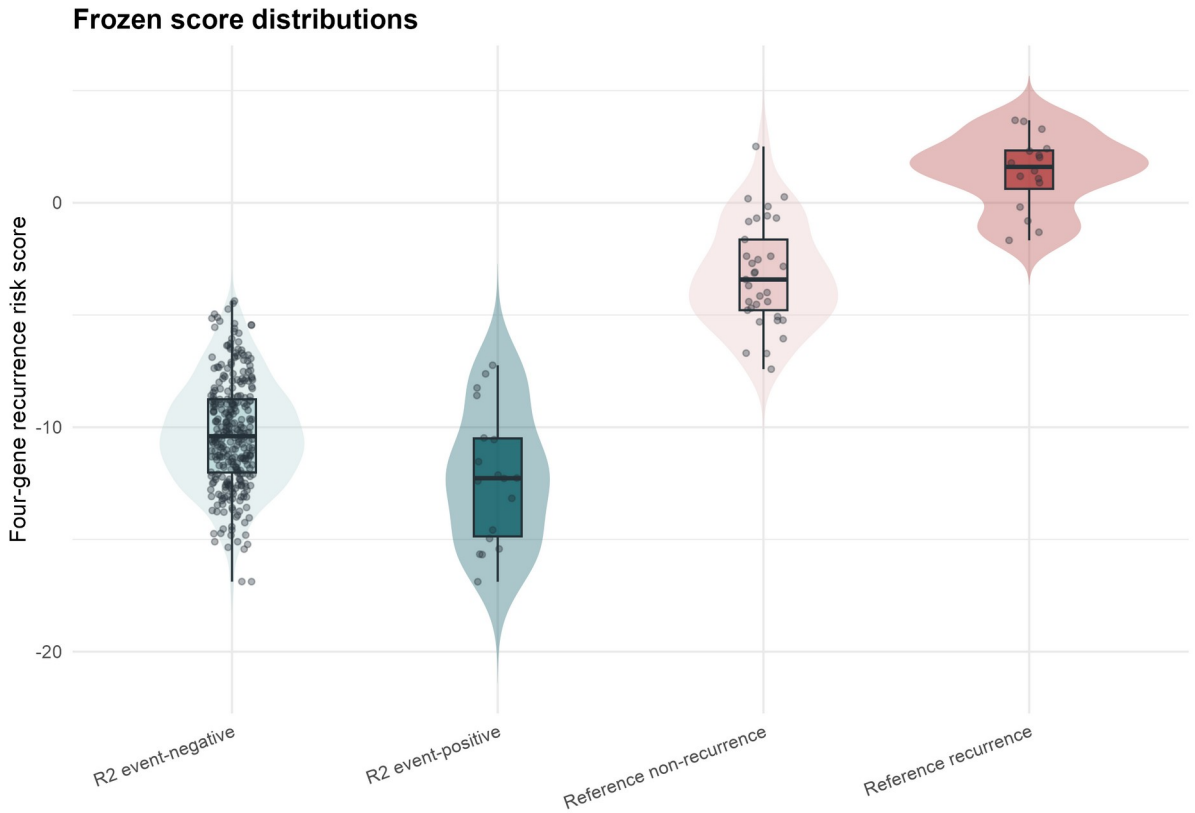
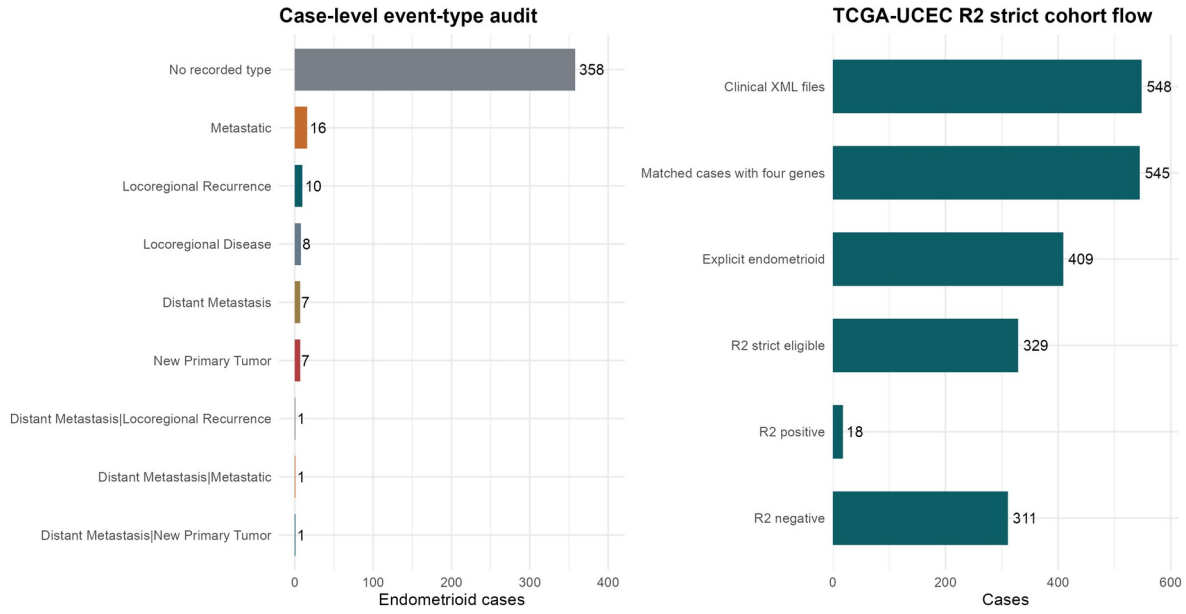


Figure 4. TCGA-UCEC event-type audit and R2 strict cohort flow. The left panel shows mutually exclusive case-level event-type combinations after XML aggregation; the right panel shows the clinical-file, expression-match, histology and R2 endpoint flow.



Supplementary files planned for submission

Additional file 1: Data-acquisition record and source roles.

Additional file 2: GSE178671 reference score table.

Additional file 3: Aggregated TCGA-UCEC field-audit summary and endpoint-adjudication table; raw per-case public-ID rows remain in the local reproducibility archive unless the journal specifically requests them.

Additional file 4: Reproducible analysis, TCGA event-type audit, and figure-generation scripts (`analyze_external_validation.py`, `audit_tcga_event_types.py`, and `build_r2_figures.R`).

Figure files: `figure1-r2-cohort-roles`, `figure2-r2-performance`, `figure3-score-distributions`, and `figure4-tcga-endpoint-audit` in PDF and PNG formats.