

Supplementary Material

for “A collective capability boundary in frontier large language models on guideline-conformant and case-specific oncology decision-making”

Sheng Zhang^{1,6,*} Jinming Li^{2,3,*} Wangyang Chen^{4,*} Zhiwei Bao^{5,†} YoSean Wang^{1,†}

¹ Institute of Digital Medicine, City University of Hong Kong, Hong Kong SAR, China

² Department of Medical Oncology, Fudan University Shanghai Cancer Center, Shanghai, China

³ Department of Colorectal Surgery, Fudan University Shanghai Cancer Center, Shanghai, China

⁴ Department of Gastroenterology, Affiliated Hangzhou First People’s Hospital, Westlake University School of Medicine, Hangzhou, Zhejiang, China

⁵ College of Biomedical Engineering and Instrument Science, Ministry of Education Key Laboratory of Biomedical Engineering, Zhejiang University, Hangzhou, China

⁶ Shaoxing Vocational & Technical College, Shaoxing, Zhejiang, China

* Equal contribution. † Corresponding authors: zwbao1996@zju.edu.cn; yosean4desci@gmail.com

Contents

S1: Clinician Adjudication Detail	1
S2: Statistical Analysis Detail	4
S3: Failure Label Mechanisms and Distribution	5
S4: CRC Case Source Index	10
S5: Custom Drug-Class Mapping	26
S6: Reproducibility	27
S7: Gold-Answer Audit on the Zero-Correct Subset	29

Supplementary S1: Clinician Adjudication Detail

This appendix documents the first-wave clinician adjudication referenced in the Methods (Clinician adjudication) and Results (Clinician validation) of the main manuscript.

S1.1 Sampling design

The 225-item adjudication sample was drawn under a stratified random sampling scheme with a frozen seed (seed = 42). Three strata were combined.

- **NCCN question-type stratum (135 items):** proportional allocation across the six question types: 77 parallel option disambiguation, 30 missing information request, 11 in-guide handoff resolution, 9 upstream routing decision, 5 evidence resolution request, 3 single-path decision. The allocation preserves the natural question-type distribution of the NCCN track while guaranteeing a minimum cell count for inter-stratum comparison.
- **Non-gastrointestinal workspace stratum (60 items):** weighted across the 59 non-GI NCCN cancer-type workspaces (breast, lung, lymphoma, melanoma, etc.). This stratum prevents over-representation of GI cancers which would otherwise dominate purely random sampling on the structured track.
- **CRC case stratum (30 items):** simple random sampling from the 419 CRC case-based items.

Items were further balanced across the nine evaluated models, with each model contributing 32–33 items. The sampling manifest, including the frozen seed, per-item assignment table, and stratum-level inclusion counts, is released with the paper artifacts (see Data Availability statement).

S1.2 Reviewer assignment and overlap design

Two board-certified oncologists each reviewed 125 items. Within each item, 25 items were assigned to *both* reviewers to support inter-rater reliability estimation ($125 + 125 - 25 = 225$ unique items). The 25-item overlap preserved the same stratification structure as the parent sample: 13 parallel option, 6 missing information, 2 in-guide handoff, 2 upstream routing, 1 evidence resolution, 1 single-path.

Reviewer assignment within each stratum was randomized using the same frozen seed. The two reviewers worked independently and were not exposed to each other’s responses during the adjudication window.

S1.3 Adjudication rubric

Each item received one of seven verdict labels:

Label	Meaning	Encoded ordinal
agree_correct	Scorer’s correct verdict is clinically appropriate	2
agree_partial	Scorer’s partial verdict is clinically appropriate	1
agree_incorrect	Scorer’s incorrect verdict is clinically appropriate	0
disagree_should_be_correct	Reviewer says output is correct, scorer wrong	2
disagree_should_be_partial	Reviewer says output is partial, scorer wrong	1
disagree_should_be_incorrect	Reviewer says output is incorrect, scorer wrong	0
unsure	Insufficient clinical context to adjudicate	—

The reviewer’s verdict (**correct** / **partial** / **incorrect**) was always recoverable from the verdict label. Free-text comments were optional and used by reviewers in roughly 40% of items, primarily to clarify the local reason for a disagreement decision.

S1.4 Headline results

The clinician adjudication sample contains 225 unique items reviewed by two oncologists, with 25 items reviewed by both for inter-rater reliability (250 reviewer–item pairs in total). After excluding the 5 pairs marked **unsure** (2 by Doctor A, 3 by Doctor B), 245 evaluable pairs remained (220 unique items).

Reviewer	Pairs evaluated	Scorer–clinician agreement	Weighted κ	95% bootstrap CI
Doctor A	123 (excl. 2 unsure)	119 / 123 (96.7%)	0.939	[0.870, 0.985]
Doctor B	122 (excl. 3 unsure)	109 / 122 (89.3%)	0.790	[0.671, 0.892]
Combined	245 evaluable pairs (220 unique items)	228 / 245 (93.1%)	—	—

Per-verdict agreement decomposition: **correct** items 95.3% (Doctor A) and 85.0% (Doctor B); **partial** items 97.3% and 91.0%; **incorrect** items 100% and 100%. Notably, all nine **incorrect**-verdict items in the adjudication sample received unanimous clinician endorsement; this directly supports the interpretation that the 42% capability boundary reflects genuine LLM failure rather than scorer false positives.

S1.5 Confusion matrices

The scorer–clinician confusion matrices for the two reviewers (rows: scorer verdict; columns: clinician-implied verdict) are reproduced below. Diagonal cells (bold) are the agreement counts; off-diagonal cells are the disagreements.

Doctor A vs. scorer ($n = 123$):

	Doctor: incorrect	Doctor: partial	Doctor: correct
Scorer: incorrect	5	0	0
Scorer: partial	0	73	2
Scorer: correct	0	2	41

Doctor B vs. scorer ($n = 122$):

	Doctor: incorrect	Doctor: partial	Doctor: correct
Scorer: incorrect	4	0	0
Scorer: partial	0	71	7
Scorer: correct	0	6	34

The two matrices share three structural features: (i) every off-diagonal cell on the **incorrect** row or column is zero, meaning no item was upgraded from or downgraded to the **incorrect** verdict by either reviewer; (ii) all disagreements fall on the **partial** $\leftrightarrow$ **correct** soft boundary; (iii) the disagreements are nearly symmetric in direction (Doctor A: 2 up, 2 down; Doctor B: 7 up, 6 down), consistent with reviewer-specific thresholds for partial credit rather than systematic scorer bias.

S1.6 Per-disagreement analysis

A total of 17 items disagreed across the 245 evaluable pairs (Doctor A: 4 items; Doctor B: 13 items). The disagreements decompose as follows:

Dimension	Distribution
Up-graded (partial $\rightarrow$ correct) vs. down-graded (correct $\rightarrow$ partial)	9 / 8 (net +1)
Involve the incorrect verdict	0 / 17
By question type	Parallel option 12, missing info 3, evidence 1, CRC 1
By model	DeepSeek V4 Pro 5, Gemini 2.5 Pro 3, Qwen 3.6-Plus 3, Minimax M2.7 2, Claude Sonnet 4.6 2, GPT-5.5 1, GLM-5 1

Twelve of seventeen disagreements (71%) come from the easiest question type (parallel option disambiguation), where model responses tend to approach the gold standard but stop short of full credit—making the **partial** / **correct** boundary inherently the most ambiguous. The harder question types (in-guide handoff resolution and upstream routing decision) yielded zero disagreements, indicating that the scorer’s judgments on hard decisions were unanimously endorsed by both clinicians.

S1.7 Inter-rater agreement on the overlap set

On the 25 overlap items, both reviewers agreed on 22 of 25 items (88.0%), yielding inter-rater Cohen’s weighted $\kappa = 0.757$ (substantial agreement, 95% bootstrap CI [0.43, 1.00], $n = 2,000$ iterations, seed = 42).

All three disagreements followed the same pattern: the scorer assigned **partial**; Doctor A agreed with **partial**; Doctor B upgraded to **correct**. The pattern is a single-direction shift consistent with Doctor B applying a slightly more lenient threshold for awarding full credit on items where the LLM’s output partially matched the gold-standard pathway. We did not find any evidence of bidirectional disagreement that would suggest a systematic scorer error.

S1.8 Worked example

A representative **agree_correct** item with the structure presented to clinicians is reproduced below in abbreviated form (item identifiers and the model name have been redacted to preserve blinding; full review packets are released with the paper artifacts).

Item ID: nccn_breast_034

Question type: parallel_option_disambiguation

Workspace: breast cancer (NCCN BINV)

Patient facts: 52-year-old female, stage IIA invasive ductal carcinoma, ER+/PR+/HER2-, sentinel lymph node biopsy positive (2/3 nodes). Genomic assay not yet performed.

Gold answer: stop_missing_info; request Oncotype DX recurrence score before adjuvant systemic therapy decision.

LLM output: decision = stop_missing_info; missing_or_needed = [Oncotype DX or comparable 21-gene recurrence-score assay]; reasoning citing parallel-option logic between adjuvant chemotherapy and endocrine therapy.

Scorer verdict: correct (slot recall = 1.0; question-type match = stop).

Clinician verdict: agree_correct.

S1.9 Data release

Full review packets, anonymized reviewer responses, per-item agreement records, disagreement annotations, sampling manifest with frozen seed, and the inter-rater confusion matrix are released at the canonical archive linked from the Data Availability statement.

Supplementary S2: Statistical Analysis Detail

This appendix documents the statistical methods underlying the confidence intervals, tier separation tests, and proportion estimates reported in Methods §2.6 and throughout Results §3.

S2.1 Bootstrap confidence intervals

Ninety-five-percent confidence intervals for NCCN strict concordance and CRC class-level therapy F1@3 were obtained by item-level non-parametric bootstrap. For each of $B = 2,000$ replications, items were resampled with replacement to the original sample size ($n_{\text{NCCN}} = 1,586$; $n_{\text{CRC}} = 419$), and the relevant aggregate statistic was recomputed on the resample. The CI lower and upper bounds correspond to the 2.5th and 97.5th percentiles of the resulting bootstrap distribution (the percentile method). A frozen random seed (`seed = 42`, NumPy default RNG) was used to make all reported intervals exactly reproducible. The bootstrap operated at the item level, not the model-output level: within each resample, all nine models were scored on the same resampled item set, preserving the paired-sample structure across models.

S2.2 Paired permutation tests for tier separation

Pairwise model comparisons used paired permutation tests over per-item NCCN strict-concordance differences. For each model pair (M_i, M_j) , we computed per-item difference scores $d_k = \text{score}_k(M_i) - \text{score}_k(M_j)$, then under the null hypothesis of exchangeability between M_i and M_j , we drew $P = 10,000$ random sign-flip vectors $\sigma \in \{-1, +1\}^n$ and computed $\bar{d}^{(p)} = \frac{1}{n} \sum_k \sigma_k d_k$. The two-sided permutation p -value is $P^{-1} \sum_p \mathbb{I}[|\bar{d}^{(p)}| \geq |\bar{d}|]$. This test makes no distributional assumption and respects the matched-pairs structure of the model evaluation (the nine models were scored on identical item sets).

S2.3 Family-wise error control

To control the family-wise error rate across the eight comparisons of adjacently ranked models (ranks 1–2, 2–3, ..., 8–9), we applied a Bonferroni correction. Setting the family-wise $\alpha_{\text{FWER}} = 0.05$ and the number of comparisons $m = 8$, the per-comparison significance threshold is $\alpha = \alpha_{\text{FWER}}/m = 0.05/8 = 0.00625 \approx 0.0063$. We considered an adjacency comparison "significant" only when its permutation p -value fell below 0.0063. Non-adjacent comparisons (e.g., rank 1 vs. rank 6) were reported as descriptive statistics only and were not used to define tier boundaries.

S2.4 Pairwise adjacency p-values

The eight adjacency p -values used to define the five tiers reported in the main paper Results (Overall performance) are summarized below.

Rank pair	Model A	Model B	Permutation p
1 vs. 2	Claude Sonnet 4.6	GLM-5.1	0.005
2 vs. 3	GLM-5.1	GLM-5	> 0.05
3 vs. 4	GLM-5	Qwen 3.6-Plus	0.61
4 vs. 5	Qwen 3.6-Plus	Minimax M2.7	> 0.05
5 vs. 6	Minimax M2.7	GPT-5.5	0.13
6 vs. 7	GPT-5.5	DeepSeek V4 Pro	< 0.001
7 vs. 8	DeepSeek V4 Pro	Gemini 2.5 Pro	< 0.001
8 vs. 9	Gemini 2.5 Pro	Gemini 3.1 Pro Preview	< 0.0001

Four of the eight adjacency boundaries crossed the Bonferroni threshold (1–2, 6–7, 7–8, 8–9). The boundaries 2–3, 3–4, 4–5, 5–6 did not separate models at $\alpha = 0.0063$ and were therefore collapsed into a single statistical tier (Tier 2: the "five-model plateau" reported in the main paper Results (Overall performance)).

S2.5 Wilson score intervals for proportions

Reported proportions (e.g., the 42.1% zero-correct rate, the 2.5% both-wrong rate in the gold-answer audit) used Wilson score 95% confidence intervals [1] rather than the normal approximation, because several proportions of interest lie near 0 or 1 where the normal interval is known to under-cover. The Wilson interval for a binomial proportion $\hat{p} = x/n$ at confidence level $1 - \alpha$ is

$$\hat{p}_{\text{Wilson}} = \frac{1}{1 + z^2/n} \left(\hat{p} + \frac{z^2}{2n} \pm z \sqrt{\frac{\hat{p}(1 - \hat{p})}{n} + \frac{z^2}{4n^2}} \right),$$

with $z = z_{1-\alpha/2}$ ($z = 1.96$ for $\alpha = 0.05$).

S2.6 Cohen’s weighted kappa

Inter-rater and scorer–clinician agreement on ordinal verdicts (**incorrect** < **partial** < **correct**; encoded 0, 1, 2) used Cohen’s weighted κ [2] with linear weights $w_{ij} = 1 - |i - j|/(k - 1)$ for $k = 3$ ordinal categories. Linear (rather than quadratic) weighting was chosen so that the penalty per category step is equal: a ‘correct’–‘partial’ disagreement and a ‘partial’–‘incorrect’ disagreement are each weighted as one step, while a ‘correct’–‘incorrect’ disagreement is weighted as two steps. CIs on κ used the percentile bootstrap at the item level ($n = 2,000$, seed = 42).

S2.7 Implementation and reproducibility

All statistical analyses were conducted in Python 3.11 using NumPy 1.26 (bootstrap, permutation), SciPy 1.11 (Wilson intervals via `scipy.stats.binomtest`), and a custom implementation of Cohen’s weighted κ that handles ordinal codings with linear weights. Random seeds for every reported analysis are recorded in the corresponding analysis script. Complete reproduction scripts are released at the canonical archive linked from the Code Availability statement.

Supplementary S3: Failure Label Mechanisms and Distribution

This appendix complements the question-type breakdown in the main paper’s anatomy-of-the-wall figure (panel A; Results, Anatomy of the wall section) with a per-failure-label view across all nine evaluated models.

S3.1 The 14 failure-label taxonomy

The deterministic scorer assigns each incorrect or partial verdict one or more typed failure labels. The labels separate route-following errors from safety errors, and the within-NCCN labels are further organized by whether the safe action was to *stop* or to *proceed*.

Label	Surface	Definition
F1 unsafe overreach	NCCN	Commits to treatment or routing when the item requires stopping for missing information or evidence.
F3 premature commitment	NCCN	Chooses one branch before a parallel-option discriminator is resolved.
F4 missed decisive info	NCCN	Fails to request a decisive missing slot.
F5 false stop	NCCN	Stops although the item has sufficient information to proceed.
F6 overcautious	NCCN	Requests unnecessary information instead of following the indicated route.
F7 wrong route	NCCN	Proceeds but selects a route that does not match the gold pathway.
F10 evidence bypass	NCCN	Proceeds despite an unresolved evidence-resolution requirement.
F11 unsupported recommendation	CRC	Recommends therapy outside the accepted set while required decision information is missing.
F12 treatment intent mismatch	CRC	Uses a treatment intent outside the accepted intent set.
F13 molecular context error	CRC	Recommends therapy inconsistent with decisive molecular context.
F14 therapy set mismatch	CRC	Output therapy set does not match the accepted therapy set.
F16 contraindicated recommendation	CRC	Recommends a therapy flagged as not recommended or molecularly contraindicated.
T1 schema violation	Tech.	Parseable JSON that violates the response schema after canonicalization.
T2 unparseable JSON	Tech.	Output cannot be parsed as a JSON object for scoring.

S3.2 Hard-failure rate (F1 + F3) per model

The hard-failure rate—defined as the union of F1 (unsafe overreach) and F3 (premature commitment) over all NCCN items (denominator = 1,586)—is the most safety-relevant single-number summary. Two models occupy an aggressive zone, exhibiting rates 3–5 $\times$ higher than the seven cautious models.

Model	F1 count	F3 count	Hard-failure rate	Zone
Qwen 3.6-Plus	33	34	4.2%	Conservative
Minimax M2.7	49	29	4.9%	Conservative
GLM-5	40	43	5.2%	Conservative
GLM-5.1	47	47	5.9%	Conservative
DeepSeek V4 Pro	52	47	6.2%	Conservative
Gemini 2.5 Pro	56	48	6.6%	Conservative
Claude Sonnet 4.6	62	60	7.7%	Conservative
GPT-5.5	161	166	20.6%	Aggressive
Gemini 3.1 Pro Preview	158	173	20.9%	Aggressive

The conservative cluster spans 4.2%–7.7% with a tight ordering driven by individual F1/F3 trade-offs (e.g., Claude Sonnet 4.6 has the highest conservative-cluster F1 rate but also the highest single-step strict concordance; Qwen 3.6-Plus has the lowest F1 rate). The two aggressive-zone models cross the 20% threshold by roughly identical margins on both F1 and F3, indicating that the aggressive behavior is not idiosyncratic to a single failure mode but a coordinated shift across both unsafe-commitment categories.

S3.3 Gemini 2.5 Pro $\rightarrow$ 3.1 Pro Preview regression decomposition

The Gemini 2.5 Pro-to-3.1 Pro Preview version-to-version transition illustrates the failure-label mechanism with particular clarity. NCCN strict concordance declined from 0.230 to 0.158 (a 31% drop) across the upgrade. Disaggregating by label reveals the mechanism.

Label	Gemini 2.5 Pro	Gemini 3.1 Pro Preview	Multiplier
F1 unsafe overreach	56	158	$\times 2.8$
F3 premature commitment	48	173	$\times 3.6$
“stop—need evidence” decisions	155	4	$\times 0.03$ (–97%)

The model became measurably more decisive (collapse of stop-for-evidence behavior) and measurably less safe (F1 and F3 both rose 2.8–3.6 $\times$). The version-to-version comparison is reported in raw counts rather than rates because it most directly exposes the magnitude of the behavioral drift.

S3.4 Knowledge-to-action gap by model

For NCCN items, the knowledge-to-action gap measures the proportion of items where the model’s reasoning text mentioned the correct next clinical step but its final decision did not reflect that knowledge (see main paper Results, Knowledge-to-action gap section). Conservative open-weight models show the largest gaps, while the aggressive Gemini 3.1 Pro Preview shows the smallest gap—because it commits more readily, often to incorrect actions.

Model	Knowledge-to-action gap
Qwen 3.6-Plus	9.1%
GLM-5	8.4%
Minimax M2.7	8.3%
Gemini 3.1 Pro Preview	3.1%

The full per-model breakdown is reported in the main paper’s anatomy-of-the-wall figure (panel C).

S3.5 Question-type breakdown

The full 9 models $\times$ 6 question types strict-concordance matrix is rendered as a heatmap in the main paper’s anatomy-of-the-wall figure (panel A). In-guide handoff resolution (IGHR) and upstream routing decision (URD) columns show uniformly low scores across all nine models, indicating that cross-pathway navigation is a class-wide blind spot rather than a single-model artifact (see main paper Results, Anatomy of the wall).

S3.6 Full 9×14 distribution

The complete 9 models $\times$ 14 failure labels count matrix is reproduced below. Counts are over the full 2,005-item evaluation set. F1–F7 and F10 are NCCN labels (denominator = 1,586); F11–F16 are CRC labels (denominator = 419); T1 and T2 are technical (parsing) labels (denominator = 2,005).

Model	NCCN labels							CRC labels					Technical	
	F1	F3	F4	F5	F6	F7	F10	F11	F12	F13	F14	F16	T1	T2
Claude Sonnet 4.6	62	60	666	39	39	124	9	9	171	1	387	25	2	0
GLM-5.1	47	47	729	48	48	163	4	36	192	2	368	21	2	0
GLM-5	40	43	694	51	51	148	3	22	217	1	384	13	1	5
Qwen 3.6-Plus	33	34	744	44	44	136	1	33	136	1	395	18	13	0
Minimax M2.7	49	29	642	38	38	133	6	25	190	2	380	20	1	48
GPT-5.5	161	166	630	24	24	154	27	22	201	1	405	14	1	0
DeepSeek V4 Pro	52	47	812	60	60	133	8	20	171	0	394	11	1	11
Gemini 2.5 Pro	56	48	892	68	68	144	6	15	177	1	373	12	3	0
Gemini 3.1 Pro Preview	158	173	940	65	65	130	24	38	103	1	390	14	0	0

Table 1: *

Counts. The same matrix expressed as per-model rates (using the track-specific denominator above) follows. Note: F5 and F6 columns coincide because the current scorer co-emits the pair (see S3.7).

Model	NCCN (% of 1,586)							CRC (% of 419)					Tech. (% of 2,005)	
	F1	F3	F4	F5	F6	F7	F10	F11	F12	F13	F14	F16	T1	T2
Claude Sonnet 4.6	3.9	3.8	42.0	2.5	2.5	7.8	0.6	2.1	40.8	0.2	92.4	6.0	0.10	0.00
GLM-5.1	3.0	3.0	46.0	3.0	3.0	10.3	0.3	8.6	45.8	0.5	87.8	5.0	0.10	0.00
GLM-5	2.5	2.7	43.8	3.2	3.2	9.3	0.2	5.3	51.8	0.2	91.6	3.1	0.05	0.25
Qwen 3.6-Plus	2.1	2.1	46.9	2.8	2.8	8.6	0.1	7.9	32.5	0.2	94.3	4.3	0.65	0.00
Minimax M2.7	3.1	1.8	40.5	2.4	2.4	8.4	0.4	6.0	45.3	0.5	90.7	4.8	0.05	2.39
GPT-5.5	10.2	10.5	39.7	1.5	1.5	9.7	1.7	5.3	48.0	0.2	96.7	3.3	0.05	0.00
DeepSeek V4 Pro	3.3	3.0	51.2	3.8	3.8	8.4	0.5	4.8	40.8	0.0	94.0	2.6	0.05	0.55
Gemini 2.5 Pro	3.5	3.0	56.2	4.3	4.3	9.1	0.4	3.6	42.2	0.2	89.0	2.9	0.15	0.00
Gemini 3.1 Pro Preview	10.0	10.9	59.3	4.1	4.1	8.2	1.5	9.1	24.6	0.2	93.1	3.3	0.00	0.00

Table 2: *

Rates (% of the track-specific denominator).

Observations directly readable from the matrix:

- **F4 (missed decisive info) dominates every model:** every model’s largest single failure label is F4, with rates ranging from 39.7% (GPT-5.5) to 59.3% (Gemini 3.1 Pro Preview). This is consistent with the headline finding that the dominant failure mode is failing to recognize cross-pathway information dependencies, not making explicit wrong-route or unsafe-commitment errors.
- **F1 and F3 separate the aggressive-vs-conservative axis:** GPT-5.5 (10.2% F1, 10.5% F3) and Gemini 3.1 Pro Preview (10.0% F1, 10.9% F3) are roughly 3 $\times$ the conservative cluster (2–4% on both).
- **F14 (CRC therapy set mismatch) is uniformly high** across all nine models (87%–97%): the CRC therapy-set match is the most demanding evaluation track and is consistent with the case-report origin of CRC gold standards documented in Supplementary S7.5.
- **T1/T2 (technical-failure) rates are below 1% for eight of nine models:** the single outlier is Minimax M2.7 at 2.39% T2 (unparseable JSON), reflecting that model’s slightly weaker JSON discipline under the structured response schema.

The per-model summary CSV (NCCN strict, NCCN lenient, CRC class F1@3 with bootstrap CIs), the per-item label assignment JSONL, and the figure-generation scripts that consume both are released with the manuscript artifacts (see the Code Availability statement).

The remaining four subsections document the deterministic mechanisms that produce the failure-label assignments tabulated above. This includes the per-label operational triggers, the slot-matching algorithm that underlies NCCN stop-type scoring, the path-matching algorithm for proceed-type scoring, and worked example failure quotations.

S3.7 Per-label operational triggers and thresholds

Each failure label is assigned by a deterministic predicate over the (gold-standard, model-output) pair. The predicates below summarize the trigger condition for each label; the full executable form is in the released scorer source.

Label	Operational trigger
F1	Item question type $\in$ {missing_information_request, evidence_resolution_request} <i>and</i> model decision = proceed.
F3	Item question type = parallel_option_disambiguation <i>and</i> model decision = proceed <i>and</i> next_step is a management branch (not workup/staging).
F4	Item is a stop-type question <i>and</i> slot recall (gold-mandated decisive slots versus model-requested slots, after synonym normalization and IDF weighting) < 0.50 .
F5	Item is a proceed-type question <i>and</i> model decision = stop_missing_info <i>and</i> model requested slots do not contain any token from the gold’s accepted-route alias set.
F6	Co-emitted with F5 in the current scorer version: whenever F5 fires, F6 is also assigned. In this release the patient-fact overlap heuristic is set conservatively to always co-trigger with F5, so F5 and F6 counts coincide in the tables below. Refining F6 to apply the overlap test independently (which would make F6 a strict subset of F5) is a deliberate future-work direction.
F7	Item is a proceed-type question <i>and</i> model decision = proceed <i>and</i> token-overlap of next_step against gold accepted-route aliases < 0.35 .
F10	Item question type = evidence_resolution_request <i>and</i> model decision = proceed (regardless of next_step quality).
F11	CRC item <i>and</i> no recommended therapy in the model output intersects the gold accepted-therapy set <i>and</i> the gold’s required-to-decide slots overlap with the model’s reported missing_information < 0.50 .
F12	CRC item <i>and</i> model’s treatment_intent is not in the gold-derived accepted-intent set (after canonicalization of synonyms such as first_line $\leftrightarrow$ induction).
F13	CRC item <i>and</i> the model recommends a therapy that violates a deterministic molecular-conflict rule (e.g., anti-EGFR for KRAS-mutant; immunotherapy as monotherapy for MSS without specific eligibility).
F14	CRC item <i>and</i> drug-family F1@3 between model recommendation and gold accepted-therapy set < 0.67 .
F16	CRC item <i>and</i> a model-recommended therapy intersects the gold’s not_recommended list.
T1	Model output is parseable JSON but fails the track-specific JSON Schema after canonicalization.
T2	Model output is not parseable as a JSON object.

Labels F2, F8, F9, and F15 are reserved for forthcoming failure modes not yet observed at sufficient frequency to warrant operational definition (the label-numbering scheme preserves these slots so that future scorer versions can add new labels without renumbering existing ones).

S3.8 Slot-matching algorithm (NCCN stop-type items)

Slot-matching is the core operation underlying F4 (missed decisive info), F6 (overcautious), and the verdict-assignment step for all NCCN stop-type items.

Input: gold-standard slot set G (e.g., {clinical stage, molecular profile, biomarker status}); model-requested slot set M (e.g., {Oncotype DX, BRCA status, treatment history}).

Step 1 (synonym normalization): each slot string is lower-cased, tokenized, and looked up in a synonym dictionary (e.g., Oncotype DX, 21-gene assay, recurrence score assay all canonicalize to recurrence_score). Synonyms are sourced from NCI thesaurus terms and validated against the items they appear in. The synonym dictionary is released alongside the scorer.

Step 2 (token-level IDF weighting): each slot is decomposed into tokens; tokens are weighted by inverse document frequency across the full NCCN gold corpus. Common tokens (e.g., status, information, history) receive low weight; clinically discriminative tokens (e.g., KRAS, Oncotype, ECOG) receive high weight.

Step 3 (slot-level F1 with IDF weighting): precision, recall, and F1 are computed at the slot level, with each slot’s contribution weighted by the maximum IDF score among its tokens. Precision-recall asymmetry: recall is the primary safety-relevant quantity (missing a decisive slot is unsafe); precision is reported for symmetry.

Step 4 (threshold to verdict): an item receives the **correct** verdict when slot recall ≥ 0.67 and content alignment score ≥ 0.50 . **Partial** is assigned when slot recall $\in [0.34, 0.67)$. **Incorrect** is assigned otherwise. The thresholds were calibrated on the pre-evaluation NCCN pilot subset and are fixed for all reported results.

S3.9 Path-matching algorithm (NCCN proceed-type items)

For proceed-type items, the scorer compares the model’s **next_step** string against the gold standard’s accepted-route alias set.

Step 1 (alias expansion): the gold accepted-route alias set is constructed from (i) the canonical route name (e.g., “adjuvant FOLFOX chemotherapy”), (ii) NCCN-recognized synonyms (e.g., “adjuvant FOLFOX”, “modified FOLFOX adjuvant therapy”), and (iii) abbreviation expansions sourced from the NCCN drug-class dictionary.

Step 2 (token overlap): the model’s **next_step** is tokenized after synonym normalization. Token overlap (Jaccard-style intersection-over-union) is computed against each alias and the maximum is retained.

Step 3 (polarity check): the model output is screened for polarity conflicts that token overlap alone would miss. Specifically, a route token sequence that matches the gold but with prefix tokens such as “palliative” instead of “adjuvant”, or “do not” negating an indicated treatment, is flagged as a polarity conflict and downgrades the verdict to **incorrect** regardless of overlap score.

Step 4 (threshold to verdict): token-overlap threshold for **correct** is 0.35. Path-matching is intentionally permissive to avoid penalizing legitimate phrasing variation; safety-relevant errors are captured by the polarity check rather than by raising the overlap threshold.

S3.10 Example failure quotations (synthetic, label-illustrative)

To preserve item-level blinding and avoid releasing real model outputs in the appendix, the examples below are synthetic constructions designed to illustrate the canonical trigger for each label. Each example uses the structured response schema described in Methods §2.3.

F1 (unsafe overreach)

Gold: **stop_missing_info** (decisive slot: clinical stage); *Model:* **proceed** with “initiate first-line FOLFOX chemotherapy”.

F3 (premature commitment)

Gold: **stop** for biomarker discriminator on parallel-option item; *Model:* **proceed** with one of the two parallel options without requesting the discriminator.

F4 (missed decisive info)

Gold missing slots: {clinical stage, BRAF status}; *Model* missing_or_needed: {appointment history, prior antibiotic exposure}.

F5 (false stop)

Gold: **proceed** on single-path item; *Model:* **stop_missing_info** requesting {baseline ECG} where the case states ECG was already obtained.

F7 (wrong route)

Gold route: “curative surgical resection”; *Model* next_step: “palliative radiation therapy”. (Polarity conflict on intent.)

F10 (evidence bypass)

Gold: **stop** for pathology confirmation on evidence-resolution item; *Model:* **proceed** citing imaging findings only.

F11 (CRC unsupported recommendation)

Gold accepted: {FOLFOX, CAPOX}; *Model:* {regorafenib monotherapy} with empty missing_information.

F13 (CRC molecular context error)

Case: KRAS-mutant, MSS; *Model:* cetuximab + chemotherapy. (Deterministic molecular-conflict rule.)

F16 (CRC contraindicated)

Gold not_recommended: {cetuximab} (KRAS-mutant); *Model* recommended: {cetuximab}.

Supplementary S4: CRC Case Source Index

This appendix lists the complete source provenance of all 419 colorectal cancer (CRC) case-based items used in the ODBB evaluation. Each case is identified by its PubMed identifier (PMID) and is freely retrievable at <https://pubmed.ncbi.nlm.nih.gov/<PMID>>.

PMID	Title	Journal	Year
32813205	Simultaneous Sigmoid Colectomy, Bilobar Liver Resection and Lung Metastasectomy via a Transdiaphragmatic Ap...	Annals of surgical oncology	2020
33413267	Severe hypercalcemia caused by parathyroid hormone in a rectal cancer metastasis: a case report	BMC Endocrine Disorders	2021
33457440	A Rare Case of Metastatic Colon Cancer in a Patient With Squamous Cell Carcinoma of the Tongue	ACG Case Reports Journal	2021
33497998	Long-term survival after surgical resection of metachronous lung, brain and thyroid gland metastases from r...	International Journal of Surgery Case Reports	2021
33563207	Cervical lymph node enlargement as the initial manifestation of rectal cancer	BMC Gastroenterology	2021
33579738	Adrenal gland as a sanctuary site for immunotherapy in patients with microsatellite instability-high metast...	Journal for Immunotherapy of Cancer	2021
33585120	Signet-Ring Cell Colon Cancer in a Teenager: A Case Report	Cureus	2021
33603403	A Case Report: Cutaneous Metastasis of Advanced Rectal Cancer with BRAF Mutation	OncoTargets and therapy	2021
33654584	Extramedullary Hematopoiesis in Mismatch Repair Deficient Colon Cancer Patient on Adjuvant Chemotherapy	Cureus	2021
33654599	Anti-EGFR Antibody Plus Chemotherapy Treatment in a Patient with Synchronous Merkel Cell Carcinoma and Colo...	Cureus	2021
33741057	Effective chemotherapy and targeted therapy supplemented with stereotactic radiotherapy of a patient with m...	Journal of Medical Case Reports	2021
33774442	Two hepatic cavernous hemangiomas mimicking colorectal liver metastasis: A rare case report	International Journal of Surgery Case Reports	2021
33857061	Metastatic Mixed Adeno-Neuroendocrine Carcinoma of the Colon to the Liver with Multiple Peritoneal Deposits...	The American Journal of Case Reports	2021
33880034	Complete Pathologic Response After Concurrent Treatment with Pembrolizumab and Radiotherapy in Metastatic C...	OncoTargets and therapy	2021
33887836	Solitary thyroid metastasis from colorectal cancer: Case report	International Journal of Surgery Case Reports	2021
33889542	Case Report: Cetuximab in Combination With Chemotherapy for the Treatment of Multifocal Hepatic Metastases ...	Frontiers in Oncology	2021
33906639	Salvage liver transplantation after resection of colorectal cancer liver metastasis with favorable outcomes...	BMC Gastroenterology	2021
33948189	Case Report—Staged brachytherapy achieving complete metabolic response in unresectable oligometastatic colo...	Oxford Medical Case Reports	2021
33974618	Colorectal Schistosomiasis Infection After Cytoreductive Surgery and Hyperthermic Intraperitoneal Chemother...	The American Journal of Case Reports	2021
33987088	Metastatic Colorectal Cancer Patient With Microsatellite Stability and BRAFV600E Mutation Showed a Complete...	Frontiers in Oncology	2021
33996527	Unexplained anemia: A case report of metachronous adenocarcinoma arising in the transverse colon following ...	Qatar Medical Journal	2021
34007446	A case of rectal cancers in teenager: A conundrum of genetics and clinical medicine	Annals of Medicine and Surgery	2021
34012676	Colorectal cancer in pregnancy: a case report and literature review	Journal of Gastrointestinal Oncology	2021
34046468	Metachronous pulmonary and pancreatic metastases arising from sigmoid colon cancer: A case report	World Journal of Clinical Cases	2021
34092783	A Case of Stage II Ascending Colon Cancer with Cardiac Tamponade Due to Pericardial Metastasis	The American Journal of Case Reports	2021

PMID	Title	Journal	Year
34159041	Simultaneous Totally Laparoscopic Distal Gastrectomy and Anterior Resection for Synchronous Gastric and Col...	Cureus	2021
34267639	A Rare BRAF Fusion in Advanced Rectal Cancer Treated with Anti-Epidermal Growth Factor Receptor Therapy	Case Reports in Oncology	2021
34268045	Capecitabine-Associated Loss of Fingerprints: A Case Report of a 62-Year-Old Man With Colorectal Cancer Suf...	Cureus	2021
34277827	Intensive multidisciplinary treatment strategies and patient resilience to challenge long-term survival in ...	Annals of Translational Medicine	2021
34290508	Complete Pathologic Response of Multiple Liver Metastases and Clinical Complete Response of Rectal Cancer i...	OncoTargets and therapy	2021
34308691	Colorectal cancer with first known case of simultaneous bilateral tibial metastases	The Journal of International Medical Research	2021
34336295	Chest Pain from Pneumopericardium with Gastropericardial Fistula	Case Reports in Cardiology	2021
34368339	Fruquintinib beneficial in elderly patient with neoplastic pericardial effusion from rectal cancer: A case ...	World Journal of Clinical Cases	2021
34380562	When a rare simultaneous infection simulates a local extension of a rectal cancer: a case report	Journal of Medical Case Reports	2021
34398037	Disseminated nocardiosis caused by Nocardia farcinica in a patient with colon cancer: A case report and lit...	Medicine	2021
34414815	Rare Distal Metastasis of Endometrial Adenocarcinoma Impersonating as Primary Colon Cancer	Journal of Investigative Medicine High Imp...	2021
34431576	Complete Pathological Response After Neoadjuvant Short-Course Immunotherapy with Ipilimumab and Nivolumab i...	The Oncologist	2021
34479591	Pedunculated early colorectal cancer with nodal metastasis: a case report	World Journal of Surgical Oncology	2021
34506464	Tumor to Tumor Metastasis: A Case Report	JNMA: Journal of the Nepal Medical Associa...	2021
34532140	A HER2-mutant patient with late-stage duodenal adenocarcinoma benefited from anti-HER2 therapy and PD-1 inh...	Journal of Gastrointestinal Oncology	2021
34535493	Rare case of metastatic adenocarcinoma to the maxillary sinus	BMJ Case Reports	2021
34540965	Genetic mutations associated with sensitivity to neoadjuvant chemotherapy in metastatic colon cancer: A cas...	World Journal of Clinical Cases	2021
34598716	Uterine cervix metastasis from primary colon adenocarcinoma: a case report and review of the literature	Journal of Medical Case Reports	2021
34600484	Genomic profiling of a patient with quadruple synchronous colorectal cancer: a case report	BMC Gastroenterology	2021
34603888	Concurrent Chemoradiation With 5-Fluorouracil and Mitomycin in Squamous Cell Carcinoma of the Rectum	Cureus	2021
34621844	Delayed diagnosis of ascending colon mucinous adenocarcinoma with local abscess as primary manifestation: R...	World Journal of Clinical Cases	2021
34622834	Rectal mixed adenoneuroendocrine carcinoma: Case report	Medicine	2021
34650810	Complete response of MSI-high metastatic colon cancer following treatment with regorafenib: A case report	Molecular and Clinical Oncology	2021
34650865	Acute Palliative Physical Therapy Services for a Patient With Metastatic Rectal Cancer and Subsequent Spina...	Cureus	2021
34665361	Lateral Nodal Features on Restaging Magnetic Resonance Imaging Associated With Lateral Local Recurrence in ...	JAMA Surgery	2019
34688262	Long-term survival of a patient with microsatellite-stable refractory colorectal cancer with regorafenib an...	BMC Gastroenterology	2021
34703908	Can adjuvant pelvic radiation therapy after local excision or polypectomy for T1 and T2 rectal cancer offer...	Clinical and Translational Radiation Oncology	2021
34786410	Diagnosis and treatment of primary pulmonary enteric adenocarcinoma: Report of Six cases	World Journal of Clinical Cases	2021
34877331	Isolated synchronous Virchow lymph node metastasis of sigmoid cancer: A case report	World Journal of Clinical Cases	2021

PMID	Title	Journal	Year
34899242	Pathological Complete Response of Clinical T4b Ascending Colon Cancer after Preoperative Chemotherapy Using...	Case Reports in Oncology	2021
34912674	Inferior Part of Rectus Abdominis Muscle Flap Outcomes after Abdominoperineal Resection: A Case Series Pilo...	World Journal of Plastic Surgery	2021
34956768	Metastatic Endometrial Cancer to the Sigmoid Colon Masquerading as Primary Colorectal Cancer	Cureus	2021
34988880	Coexistence of gastrointestinal stromal tumour and colorectal adenocarcinoma: Two case reports	Journal of Gastrointestinal Oncology	2011
35003463	A Rare Case of Colon Cancer Metastasis to the Brain anda Brief Review of its Treatment and Prognosis	Radiology Case Reports	2021
35003780	Rare BRCA2 Mutation and MSI-H in Male Patient with Aggressive Stage III Colorectal Cancer, Case Report and ...	Current Health Sciences Journal	2021
35012456	Complete pathological response of colorectal peritoneal metastases in Lynch syndrome after immunotherapy ca...	BMC Gastroenterology	2022
35027849	Rectal Cancer Metastasis to the Anal Verge: An Unusual Case Presentation and Review of the Literature	International Medical Case Reports Journal	2022
35049209	A new spray-type adhesion barrier (AdSpray) improves condition for surgical treatment in the reversal of Ha...	Medicine	2021
35071093	Metastatic Colon Cancer in the Pituitary: A Case Report with Review of Literature	Asian Journal of Neurosurgery	2021
35082635	Early Detection of Circulating Tumor DNA Postoperatively Enables Discovery of Resectable Metastatic Disease...	Case Reports in Oncology	2021
35083362	A Case of Esophageal Squamous Cell Carcinoma Metastasized to the Colonic Anastomotic Site of Right Hemicolectomy	ACG Case Reports Journal	2022
35116564	Adenosquamous carcinoma of the hepatic flexure of colon: a case report	Translational Cancer Research	2021
35127928	Recurrence of sigmoid colon cancer-derived anal metastasis: A case report and review of literature	World Journal of Clinical Cases	2022
35128139	Peritoneal Recurrence of Cecal Cancer with Specific Imaging Findings and Shrinkage after Treatment with Pem...	Journal of the Anus, Rectum and Colon	2022
35129794	Carcinosarcoma of the stomach: A rare tumor for an unusual localization. Review of the literature	The Turkish Journal of Gastroenterology	2019
35140577	Anterior abdominal wall metastasis following curative resection and chemoradiation of rectal cancer masquer...	Journal of Taibah University Medical Sciences	2021
35154775	Effective management of advanced colon cancer genotyping microsatellite stable/microsatellite instable-low ...	SAGE Open Medical Case Reports	2021
35172772	Synchronous “skip” facial metastases from colorectal adenocarcinoma: a case report and review of literature	BMC Gastroenterology	2022
35178262	Colonic Adenocarcinoma with Plasmacytoid Feature: Histopathology and Molecular Characteristics of a Rare Ne...	Case Reports in Pathology	2022
35186253	Locally advanced rectal cancer in a patient with ectopic kidney: case report and literature review	Journal of Surgical Case Reports	2022
35186512	Improvement of Metastatic Spinal Cord Compression After Decompression Surgery and Radiotherapy in a Patient...	Cureus	2022
35198145	Complete pathological response of post neoadjuvant chemotherapy in colorectal cancer with liver metastasis:...	Journal of Surgical Case Reports	2022
35226656	Delayed Diagnosis of Clostridium difficile Colitis in a 48-Year-Old Woman with a Homozygous Mutation of the...	The American Journal of Case Reports	2022
35233281	Breast metastasis from rectal carcinoma: A case report and review of the literature	Malawi Medical Journal	2021
35251973	Complete Response With Cetuximab-Based Treatment of Metastatic Colorectal Cancers: Two Case Reports and Lit...	Frontiers in Oncology	2022
35296259	Combined laparoscopic lymphadenectomy of lateral pelvic and inguinal nodal metastases using indocyanine gr...	BMC Gastroenterology	2022

PMID	Title	Journal	Year
35321695	Giant ovarian tumor with colorectal cancer: suggestion concerning the need for colonoscopy screening in cas...	BMC Surgery	2022
35399448	Importance of Early Next-Generation Sequencing in Microsatellite Unstable Colon Cancer With a High Tumor Mu...	Cureus	2022
35433988	Efficacy of sintilimab and fruquintinib combination treatment in the management of microsatellite-stable me...	Annals of Translational Medicine	2022
35464894	Sacral osteomyelitis as a rare cause of anorectal pain several years following treatment for rectal carcinoma	Oxford Medical Case Reports	2022
35571658	Gene profiling of metastatic small intestinal squamous cell carcinoma after lung squamous cell carcinoma su...	Translational Cancer Research	2022
35572475	An Extremely Rare Case of Rectal Signet Ring Cell Carcinoma	Gastroenterology Research	2022
35611253	Therapeutic impact of determination of RAS mutations in the plasma of patient with colorectal cancer	Gastroenterology and Hepatology From Bed t...	2022
35619103	Locally advanced rectal cancer receiving total neoadjuvant therapy combined with nivolumab: a case report a...	World Journal of Surgical Oncology	2022
35623078	Concurrent ascending colon adenocarcinoma and ileocecal tuberculosis: A case report	Medicine	2022
35628872	Electrochemotherapy of Primary Colon Rectum Cancer and Local Recurrence: Case Report and Prospective Analysis	Journal of Clinical Medicine	2022
35659714	Oligometastatic deposits of prostate cancer found within the sigmoid pericolic fat that was resected for co...	Journal of Medical Case Reports	2022
35663075	Primary sigmoid squamous cell carcinoma with liver metastasis: A case report	World Journal of Clinical Cases	2022
35663084	PD-1 inhibitor in combination with fruquintinib therapy for initial unresectable colorectal cancer: A case ...	World Journal of Clinical Cases	2022
35720492	A pathological complete response after nivolumab plus ipilimumab therapy for DNA mismatch repair-deficient/...	Oncology Letters	2022
35734602	KRAS Gene Copy Number as a Negative Predictive Biomarker for the Treatment of Metastatic Rectal Cancer With...	Frontiers in Oncology	2022
35734715	Rectal metastasis originating from breast cancer: A rare case report	Annals of Medicine and Surgery	2022
35761909	Organ Preservation in Colon Cancer: An Illustrative Case Report	Cureus	2022
35815302	Larotrectinib in Mismatch-Repair-Deficient TRK Fusion-Positive Metastatic Colon Cancer After Progression on...	Cureus	2022
35837187	Immunotherapy in synchronous MSI-H rectal adenocarcinoma and upper tract urothelial carcinoma: a case report	Journal of Gastrointestinal Oncology	2022
35871578	Metastatic Lung Tumors from Colorectal Cancer with EGFR Mutations That Responded to Osimertinib	Internal Medicine	2022
35891856	Guillain-Barré-Like Syndrome From Esophageal Squamous Cell Carcinoma	Cureus	2022
35951216	Tumor-to-tumor metastasis of colon cancer metastasizing to a pancreatic neuroendocrine tumor associated wit...	Clinical Journal of Gastroenterology	2022
35979324	Transplanted kidney loss during colorectal cancer chemotherapy: A case report	World Journal of Clinical Cases	2022
36035642	Colon Cancer Metastasis to the Right Testis: Case Report and Review of Literature	Case Reports in Urology	2022
36046437	Lynch syndrome-associated lung cancer: pitfalls of an immunotherapy-based treatment strategy in an unusual ...	Exploration of Targeted Anti-tumor Therapy	2021
36059435	Immunotherapy for metastatic liver disease from colorectal carcinoma: case series from the Middle East	Journal of Surgical Case Reports	2022
36086719	Iodine-125 seed implantation combined chemotherapy for metastatic pancreatic adenocarcinoma with primary co...	Medicine	2022
36118993	Role of dual anti-HER2 therapy before liver resection for rectal metastases	Journal of Surgical Case Reports	2022

PMID	Title	Journal	Year
36119063	Case report: The MSI-L/p-MMR metastatic rectal cancer patient who failed systemic therapy responds to anti-...	Frontiers in Immunology	2022
36192748	Intramural cancer recurrence in the rectum after curative surgery for proximal sigmoid colon cancer: a case...	World Journal of Surgical Oncology	2022
36252546	An unusual delayed rectal metastasis from prostate cancer masquerading as primary rectal cancer	International Journal of Surgery Case Reports	2022
36253824	Literature review of imaging, pathological diagnosis, and outcomes of metachronous lung and pancreatic meta...	World Journal of Surgical Oncology	2022
36254007	Platelet-rich plasma for the treatment of chronic rectal ulcer: A case report	Medicine	2022
36348904	A Rare Case of Bilateral Orbital Metastases From Rectal Adenocarcinoma	Cureus	2022
36353566	Case Report: A case of advanced duodenal adenocarcinoma in complete remission after chemotherapy combined w...	Frontiers in Oncology	2022
36387791	Metachronous isolated penile metastasis from sigmoid colon adenocarcinoma: A case report	World Journal of Clinical Cases	2022
36388049	A case report of pulmonary combined small cell carcinoma with enteric adenocarcinoma	Translational Cancer Research	2022
36388681	Conversion surgery by hand-assisted laparoscopic surgery following chemotherapy for rectal cancer with H3 l...	Journal of Gastrointestinal Oncology	2022
36388683	Preliminary experience of oral fruquintinib-capecitabine as a new maintenance treatment strategy for advanc...	Journal of Gastrointestinal Oncology	2022
36388686	Three-target treatment combined with surgery for BRAF V600E-mutant colon cancer with peritoneal metastasis:...	Journal of Gastrointestinal Oncology	2022
36404497	Peritoneal metastatic mixed adenoneuroendocrine carcinoma treated with cytoreductive surgery and hypertherm...	Annals of Coloproctology	2022
36438698	Silent advanced large cell neuroendocrine carcinoma with synchronous adenocarcinoma of the colon: A case re...	World Journal of Gastrointestinal Oncology	2022
36451153	Breast metastasis of signet ring cell carcinoma from the colon: a case report	World Journal of Surgical Oncology	2022
36457540	A case of gastric metastasis originating from right-sided colon cancer 4 years after colectomy	Future Science OA	2022
36460309	Linitis plastica of the rectum secondary to prostate carcinoma	BMJ Case Reports	2022
36483826	T1 rectal mucinous adenocarcinoma with bilateral enlarged lateral lymph nodes and unilateral metastasis: A ...	World Journal of Clinical Cases	2022
36518767	Case report: Complete response to pembrolizumab in a liver metastatic colon adenocarcinoma patient with a n...	Frontiers in Immunology	2022
36531025	Anlotinib combined with TAS-102 as the third-line treatment for a patient with metastatic colon cancer: A c...	Frontiers in Oncology	2022
36544760	Pathological complete response after neoadjuvant immunotherapy combined with chemotherapy in pediatric rect...	Frontiers in Immunology	2022
36591529	Post-operative ctDNA monitoring in stage I colon cancer: A case report	Frontiers in Oncology	2022
36605844	Colorectal liver metastasis with bile duct tumor thrombus discovered 15 years post primary tumor resection:...	International Cancer Conference Journal	2022
36615039	Pathological Complete Response Following Neoadjuvant Tislelizumab Monotherapy in Treatment-Naive Locally Ad...	Journal of Clinical Medicine	2022
36647049	Diagnostic and therapeutic challenges of glioblastoma as an initial malignancy of constitutional mismatch r...	BMC Medical Genomics	2023
36742094	A Patient with Transverse Colon Cancer Complicated by Cowden Syndrome Administered FOLFOXIRI + Bevacizumab ...	Case Reports in Gastroenterology	2023
36763235	Clinical utility of chromogranin A and octreotide in large cell neuro endocrine carcinoma of the uterine co...	Rare Tumors	2011
36772901	Regorafenib-Induced Radiation Recall Presenting as Acute Blood Loss Anemia With Rectal Bleeding and Severe ...	Journal of Investigative Medicine High Imp...	2023

PMID	Title	Journal	Year
36776321	Long-term stable disease with mFOLFOX6 chemotherapy plus cetuximab for bone marrow metastasis from rectal c...	Frontiers in Oncology	2023
36801767	Pathological complete response of multiple liver metastases from colon cancer successfully treated with con...	International Journal of Surgery Case Reports	2023
36810548	Priapism in a Patient with Rectal Adenocarcinoma	Diseases	2023
36814823	Heterogeneous disease and intermittent treatment in metastatic colorectal cancer: A case report	Frontiers in Oncology	2023
36816936	Case report: Preclinical efficacy of NEDD8 and proteasome inhibitors in patient-derived models of signet ri...	Frontiers in Oncology	2023
37020543	Abscopal effect triggered by radiation sequential mono-immunotherapy resulted in a complete remission of PM...	Frontiers in Immunology	2023
37033359	Unusual presentation of a neuroendocrine tumor in the ileostomy specimen after rectal cancer treatment: a c...	Translational Cancer Research	2023
37034492	Metastasis of colorectal adenocarcinoma to the right ventricle in a young man: a case report	The Pan African Medical Journal	2023
37064498	Circulating Tumor DNA Testing Overcomes Limitations of Comprehensive Genomic Profiling from Tumor Tissue	Case Reports in Oncology	2023
37114121	Case report: anaplastic lymphoma kinase (ALK) rearranged adenocarcinoma with high level of microsatellite i...	Frontiers in Oncology	2023
37114157	Rectal prolapse as the initial presentation of rectal cancer—A case report	Frontiers in Surgery	2023
37123317	Successful treatment of breast metastasis from primary transverse colon cancer: A case report	World Journal of Clinical Cases	2023
37129184	Cetuximab plus XELOX show efficacy against brain metastasis from colorectal cancer: a case report	CNS Oncology	2023
37146556	Challenges presented by complete response to immune checkpoint blockade in patients with dMMR colorectal ca...	International Journal of Surgery Case Reports	2023
37164671	Cutaneous Metastasis of Transverse Colon Cancer with an Aberrant Pattern of CK7/CK20/CDX2 and High Microsat...	Internal Medicine	2023
37188188	PD-1 inhibitor combined with radiotherapy and GM-CSF in MSS/pMMR metastatic colon cancer: a case report	Frontiers in Oncology	2023
37611397	A rare case of sigmoid intussusception. A case report	International Journal of Surgery Case Reports	2023
37621810	Successful Treatment of pMMR MSS IVB Colorectal Cancer Using Anti-VEGF and Anti-PD-1 Therapy in Combination...	Cureus	2023
37652473	Rapid Reduction of CEA and Stable Metastasis in an NRAS-mutant Rectal-Cancer Patient Treated With FOLFIRI a...	In Vivo	2023
37701877	Colon Metastasis from Pancreatic Cancer: A Case Report	OncoTargets and Therapy	2023
37731579	Collision tumor of primary malignant lymphoma and adenocarcinoma in the colon diagnosed by molecular pathol...	World Journal of Clinical Cases	2023
37815339	Rectal metastasis of gastric cancer: a case report	The Journal of International Medical Research	2023
37818131	Adaptive radiotherapy of locally advanced sigmoid colon cancer with intra-fractional motion using the MRIdi...	Oncology Letters	2023
37840105	Identification of patient subgroups with unfavorable long-term outcomes associated with laparoscopic surger...	Annals of Gastroenterological Surgery	2021
38033418	Anti-epidermal growth factor receptor treatment for patients with NeORAS wild-type metastatic colorectal ca...	Therapeutic Advances in Medical Oncology	2023
38038857	Prognostic factors and sites of metastasis in unresectable locally advanced pancreatic cancer	Cancer Medicine	2015
38046915	Signet-ring cell carcinoma of the caecum: A case report	Radiology Case Reports	2023
38065910	Resection of rectal metastasis after previous radical surgery for pancreatic cancer: Case report and litera...	Medicine	2023
38074063	Synchronous Gastric and Colon Cancer	Cureus	2023

PMID	Title	Journal	Year
38078126	Isolated axillary tumor deposit consistent with primary breast carcinoma: A case report	World Journal of Clinical Cases	2023
38091709	Challenging case of deficient mismatch repair right-sided locally advanced adenocarcinoma of the ascending ...	International Journal of Surgery Case Reports	2023
38106808	A Rare Case of Cecal Adenocarcinoma Presenting as Intussusception	Cureus	2023
38111764	Organ sparing to cure stage IV rectal cancer: A case report and review of literature	World Journal of Gastrointestinal Surgery	2023
38124787	Exceptional Response in BRAF p.V600E-Mutant Enteric-Type Adenocarcinoma of the Lung With Cutaneous Spread: ...	JTO Clinical and Research Reports	2023
38173434	Conversion immunotherapy for deficient mismatch repair locally unresectable colon cancer: A case report	World Journal of Gastrointestinal Oncology	2023
38173715	Case report: hepatic arterial infusion chemotherapy combined with sintilimab and lenvatinib for conversion ...	Frontiers in Immunology	2023
38196522	Significant response from fruquintinib plus anti-PD-1 immunotherapy for microsatellite stable metastatic co...	Journal of Gastrointestinal Oncology	2023
38229786	Stage II Medullary Carcinoma of the Colon: A Surgery Case Report	Cureus	2023
38283131	Single metachronous bone metastasis following rectal adenocarcinoma: A case report	Medicine International	2024
38283860	Long-term response with the atypical reaction to nivolumab in microsatellite stability metastatic colorecta...	Drug Target Insights	2024
38292846	Emerging role of liquid biopsy in rat sarcoma virus mutated metastatic colorectal cancer: A case report	World Journal of Gastrointestinal Oncology	2024
38299094	Synchronous Malignant Peritoneal Mesothelioma and Sigmoid Adenocarcinoma: a Challenging Clinical Entity	Medical Archives	2023
38304657	Apple Core Unveiled: Malignant Colonic Obstruction Revealing an Unknown Rectosigmoid Neoplasm With Foreign ...	Cureus	2024
38310787	Concurrent rectal perforation and obstruction following neoadjuvant chemoradiation for locally advanced rec...	International Journal of Surgery Case Reports	2024
38313923	A Rare Case of Cutaneous Metastasis of Unresectable Rectal Adenocarcinoma	Cureus	2024
38321556	Radiological complete response with regorafenib for multiple lung metastases of ascending colon cancer: a c...	Journal of Medical Case Reports	2024
38322677	Adenocarcinoma of sigmoid colon with metastasis to an ovarian mature teratoma: A case report	World Journal of Clinical Cases	2024
38335914	Liver function and image evaluation after radiotherapy for liver metastases after resection of sigmoid colo...	International Journal of Surgery Case Reports	2024
38380204	An Intrahepatic Fluorodeoxyglucose (FDG)-PET/CT False-Positive Tumor Secondary to Foreign Body Granuloma Ma...	Cureus	2024
38381603	Anaplastic Lymphoma Kinase (ALK) Inhibitors Show Activity in Colorectal Cancer With ALK Rearrangements: Cas...	The Oncologist	2024
38390428	Intrabiliary metastasis of colorectal mucinous adenocarcinoma mimicking choledocholithiasis 18 years after ...	Radiology Case Reports	2024
38413220	A case of dMMR/MSI-H/TMB-H colon cancer with brain metastasis treated with PD-1 monoclonal antibody	Journal of Zhejiang University (Medical Sc...	2024
38425406	Anti-EGFR antibody monotherapy for colorectal cancer with severe hyperbilirubinemia: A case report	World Journal of Gastrointestinal Oncology	2024
38449931	A Case Report on Alpha-Fetoprotein-Positive Colorectal Cancer	Cureus	2024
38482252	An unexpected case of large cell neuroendocrine carcinoma of the colon: a case report	Journal of Gastrointestinal Oncology	2024
38496115	A Rare Case of Duodenal Adenocarcinoma With Brain Metastasis	Cureus	2024
38500924	Radical Resection for Locally Advanced Colon Cancer With Bladder Involvement Treated in a Tertiary Health C...	Cureus	2024

PMID	Title	Journal	Year
38549624	An Improvement of Exertional Dyspnea by the Reintroduction of Anti-EGFR Antibody in Patients with Metastati...	Case Reports in Oncology	2024
38558649	Genomic Analysis of Undifferentiated Carcinoma of the Pancreas with Squamous Differentiation: A Case Report	Cureus	2024
38558747	Delayed Postoperative Intracerebral Hemorrhage Associated With Oral Multikinase Inhibitor Therapy for Cance...	Cureus	2024
38559408	Rectal adenocarcinoma in the eye: An unexpected destination	SAGE Open Medical Case Reports	2024
38586180	Panitumumab-Induced Periorbital Dermatitis: A Case Report	Clinical, Cosmetic and Investigational Der...	2024
38590724	Personalized medicine for locally advanced rectal cancer: five years of complete clinical response after ne...	Frontiers in Surgery	2024
38608071	Direct gastric invasion from the liver metastasis of colorectal origin: A case report	Medicine	2024
38617840	Complete remission in a pretreated, microsatellite-stable, KRAS-mutated colon cancer patient after treatmen...	Frontiers in Immunology	2024
38631859	Tumor Lysis Syndrome in a Patient with BRAFV600E Mutated Colon Cancer Treated with Cetuximab and Encorafenib	Internal Medicine	2024
38633978	Individualized Treatment Approach for Rectal Adenocarcinoma in the Setting of Congenital Neutropenia	Cureus	2024
38638862	Clinical outcomes of intermittent panitumumab based-therapy for previously treated older patient with metas...	Frontiers in Oncology	2024
38638920	Oligometastatic colorectal adenocarcinoma to the spleen and ovaries	Journal of Surgical Case Reports	2024
38646233	Synchronous Gastric and Colonic Adenocarcinoma: A Case Report With its Molecular Implications	Cureus	2024
38646436	Paratesticular metastasis from colorectal adenocarcinoma presenting as hydrocele: a rare case report and li...	Frontiers in Oncology	2024
38650764	Re-Irradiation With Proton Beam Therapy for Localized Perineural Spread Following Presacral Recurrence in S...	Cureus	2024
38681488	Reactive lymphoid hyperplasia of the liver after surgery for advanced sigmoid colon cancer: a case report	Journal of Surgical Case Reports	2024
38689672	Pathologic Complete Response and Long-Term Survival After Preoperative Chemotherapy for Transverse Colon Ca...	Cureus	2024
38693541	An isolated vaginal metastasis from rectal cancer: a case report	Journal of Medical Case Reports	2024
38694366	Concurrent ascending colon adenocarcinoma and colonic tuberculosis: a case report from Syria	Annals of Medicine and Surgery	2024
38711893	Partial response to crizotinib + regorafenib + PD-1 inhibitor in a metastatic BRAF V600EMT colon cancer pat...	AME Case Reports	2024
38741816	Metastasectomy and Stereotactic Body Radiotherapy for Colorectal Cancer With Liver and Lung Oligometas-tases...	Cureus	2024
38764824	Chemoradiotherapy plus tislelizumab for mismatch repair proficient rectal cancer with supraclavicular lymph...	World Journal of Gastrointestinal Oncology	2024
38774401	Rare metastasis in a patient with BRAF-mutated rectal cancer: choroidal metastasis – case report and litera...	GMS Ophthalmology Cases	2024
38821532	Colorectal cancer harboring EGFR kinase domain duplication response to EGFR tyrosine kinase inhibitors	The Oncologist	2024
38855326	Sign of Neck Mass as the Chief Complaint: A Case Report and Literature Review About Thyroid Metastasis of C...	Cancer Management and Research	2024
38868260	A Rare Case of Insulin-Like Growth Factor (IGF-2) Induced Hypoglycemia Associated With Metastatic Colon Cancer	Cureus	2024
38869444	BRAF V600E-mutant colorectal cancer with CNS metastases treated successfully with encorafenib, binimetinib ...	CNS Oncology	2024

PMID	Title	Journal	Year
38895054	Combination of Brucea javanica oil emulsion and Aidi injection associated with the long-term survival of a ...	Oncology Letters	2024
38903319	Synchronous Gastric and Rectal Adenocarcinoma With Complete Response After Total Neoadjuvant Therapy: A Cas...	Cureus	2024
38926771	Ascending colon cancer metastasized to the right testicle: a case report	Journal of Medical Case Reports	2024
38952672	Complete response of metastatic microsatellite-stable BRAF V600E colorectal cancer to first-line oxaliplati...	Oncoimmunology	2024
38978969	A Case Report of Pathologically Complete Response of a Huge Lymph Node Metastasis of Colorectal Cancer Afte...	ImmunoTargets and Therapy	2024
38988445	Combination of microwave ablation and systemic treatments achieve a long survival time for a patient with m...	Oncology Letters	2024
38988695	Long-Term Survival in a Patient with Metastatic Colorectal Cancer Treated with Trifluridine/Tipiracil as La...	Case Reports in Oncology	2024
38989442	An unusual presentation of solitary ovarian metastasis from colorectal cancer in an elderly woman: a case r...	Journal of Gastrointestinal Oncology	2024
39003365	Sessile serrated lesions with dysplasia: is it possible to nip them in the bud?	Journal of Gastroenterology	2023
39011226	An Unusual Case of Lynch Syndrome	Cureus	2024
39011472	Case report: Precision guided reactive cancer management: molecular complete response in heavily pre-treated...	Frontiers in Oncology	2024
39067098	Peno-scrotal metastasis of colorectal adenocarcinoma: A case report and review of the literature	International Journal of Surgery Case Reports	2024
39072179	BRAF K601E-mutated metastatic colorectal cancer in response to combination therapy with encorafenib, binime...	World Journal of Gastrointestinal Oncology	2024
39072183	Metachronous multifocal carcinoma: A case report	World Journal of Gastrointestinal Oncology	2024
39086877	Successfully Resected Isolated Lateral Lymph Node Recurrence in a Patient with T1 Lower Rectal Cancer: Case...	Journal of the Anus, Rectum and Colon	2024
39114842	Neoadjuvant immunotherapy leads to complete pathologic response in locally advanced colon cancer	Clinical Case Reports	2024
39135839	An Unusual Tumor: Squamous Cell Carcinoma of the Colon in Lynch Syndrome	Cureus	2024
39161673	Cardiac interventricular septum hemangioma in a colon cancer patient treated with Capecitabine: A case repo...	Clinical Case Reports	2024
39167100	Robotic Colorectal Surgery	Yonsei Medical Journal	2008
39182305	Sister Mary Josphe's nodule as metastasis of colorectal cancer. Systematic review of the literature and met...	International Journal of Surgery Case Reports	2024
39193330	Case report: Pyrotinib and tegafur combined with radiotherapy achieved notable response in HER2-amplified r...	Frontiers in Pharmacology	2024
39211303	Bicentric lesion of colon cancer with postoperative fever: A case report	Oncology Letters	2024
39241479	Concurrent rectosigmoid cancer and Krukenberg tumor in a pregnant patient: A rare clinicopathological prese...	International Journal of Surgery Case Reports	2024
39247185	Durable response to pembrolizumab in hepatic metastasis from colonic carcinoma with Lynch syndrome: a case ...	Frontiers in Immunology	2024
39266793	Anatomy of the smooth muscle structure in the female anorectal anterior wall: convergence and anterior exte...	Colorectal Disease	2019
39268170	Clinicopathological features of endometriosis-associated adenocarcinoma of the rectum: A report of two cases	Oncology Letters	2024
39279969	DNA signature and occult blood in stool led to diagnosis of an early-onset metastatic rectal cancer in a yo...	Journal of Gastrointestinal Oncology	2024
39310573	Curative Surgery After Neoadjuvant Chemotherapy for Locally Advanced Sigmoid Colon Cancer With Extensive Ab...	Cureus	2024

PMID	Title	Journal	Year
39351455	Successful cetuximab rechallenge in metastatic colorectal cancer: A case report	World Journal of Clinical Oncology	2024
39413642	A “one in a million” case of colorectal carcinoma – A case report from a tertiary care centre in Mumbai, India	International Journal of Surgery Case Reports	2024
39430753	A watch-and-wait approach for metachronous multiple colon cancer following neoadjuvant immunotherapy: a cas...	Frontiers in Immunology	2024
39435196	A Case of Oropharyngeal Carcinoma Accompanying a Presacral Malignant Epidermoid Cyst	Cureus	2024
39463904	Unusual Metastatic Pathways: A Case of Colorectal Adenocarcinoma Presenting as a Cervical Neck Mass	Cureus	2024
39473719	Formation of a Colo-colonic Fistula Communicating with the Transverse Colon in Cecal Cancer: A Case Report	Journal of the Anus, Rectum and Colon	2024
39496016	Extremely rare mucinous adenocarcinoma of the small bowel causing bilateral metastatic Kukenberg tumors of ...	Medicine	2024
39506835	Total Minimally Invasive Curative Staged Resections After Induction Systemic Therapy for Metastatic Rectal ...	Cancer Reports	2024
39552190	ctDNA-guided adjuvant immunotherapy in colorectal cancer	Immunotherapy	2024
39558953	Perianal leiomyosarcoma as a rare sequela of rectal cancer radiotherapy: a case report	Frontiers in Oncology	2024
39560550	Neoadjuvant chemo-radiotherapy combined with immune checkpoint inhibitors: A case report of rectal small-ce...	Medicine	2024
39602930	Undifferentiated spindle cell sarcoma at the anastomosis after ileocecal resection for colon cancer: A case...	International Journal of Surgery Case Reports	2024
39606043	An unexpected primary squamous cell carcinoma of the left colon: a rare case report	Journal of Surgical Case Reports	2024
39626159	Evolution of Rapid Clonal Dynamics and Non–Cross-Resistance in Response to Alternating Targeted Therapy and...	JCO Precision Oncology	2024
39687148	Rare BRAF gene fusions in metastatic early-onset colon cancer: A case report	Heliyon	2024
39687889	The use of nutrigenomics and nutritional biomarkers with standard care of long-term recurrent metastatic re...	Frontiers in Oncology	2024
39696980	Diagnostic Challenges due to a Germline Missense MSH2 Variant in a Patient With Immunotherapy-Responsive Lo...	Cancer Reports	2024
39713608	Cutaneous metastases from rectal adenocarcinoma: A case report	SAGE Open Medical Case Reports	2024
39735101	Pushing the Boundaries of Minimally Invasive Surgery: Fully Laparoscopic Left Hepatectomy Extended to Segme...	Cureus	2024
39759155	Case report: Durable response from tegafur/gimeracil/oteracil (S-1) combined with fruquintinib and sintilim...	Frontiers in Oncology	2024
39789278	Locally Recurrent Rectal Cancer: Toward a Second Chance at Cure? A Population-Based, Retrospective Cohort S...	Annals of Surgical Oncology	2023
39789783	Efficacy of Combined PD-1 Inhibitor and Bevacizumab in Unresectable Liver Metastasis of MSI-H Colorectal Ca...	The American Journal of Case Reports	2025
39803152	Laparoscopic Resection of an Isolated Myometrial Metastasis From Ileocecal Cancer: A Case Report	Cureus	2024
39811421	Durable Remission After Targeted Therapy in BRAF V600E–Mutant Metastatic Colorectal Cancer: Case Report	Journal of Immunotherapy and Precision Onc...	2025
39835129	Case report: Watch-and-wait strategy in resectable esophageal cancer following neoadjuvant chemoimmunothera...	Frontiers in Immunology	2025
39853260	Fertility-sparing treatment outcomes using immune checkpoint inhibitors in endometrial cancer patients with...	Journal of Gynecologic Oncology	2025
39866256	Differential response to immunotherapy in different lesions of MSI-H double primary colorectal cancer: a ca...	AME Case Reports	2024

PMID	Title	Journal	Year
39889652	Primary small rectal neuroendocrine tumor with pelvic lateral lymph nodes metastasis: A case report and rev...	International Journal of Surgery Case Reports	2025
39917173	Pulmonary enteric adenocarcinoma with progression disease after second - line therapy: a case report	Frontiers in Oncology	2025
39926284	Cerebellar metastasis from colorectal cancer: a case report	Frontiers in Oncology	2025
39963396	A rare case of sporadic metastatic colorectal cancer with glioblastoma multiforme: a challenging clinical s...	Journal of Surgical Case Reports	2025
39966961	Genetic analysis for diagnosing local recurrence of sigmoid colon cancer mimicking a small intestinal tumor...	World Journal of Surgical Oncology	2025
39980522	Metastatic Colonic Adenocarcinoma Mimicking Acute Bowel Perforation: Hiccups and Vomiting as Unusual Initia...	Case Reports in Oncology	2024
39991497	Spontaneous Regression of Advanced Transverse Colon Cancer: A Case Report	Surgical Case Reports	2025
39996388	Durable response to third-line combination therapy in a metastatic colorectal cancer patient with BRAF V600...	Human Vaccines & Immunotherapeutics	2025
40026534	Early Detection of Krukenberg Tumors Utilizing ctDNA Testing and CEA Monitoring	Case Reports in Oncological Medicine	2025
40029573	Spontaneous regression of transverse colon cancer with high-frequency microsatellite instability: a case re...	World Journal of Surgical Oncology	2019
40051807	Advanced rectal cancer in an 11-year-old Sudanese patient with rare metastatic site: a case report	Journal of Surgical Case Reports	2025
40092928	Penile metastasis from colon cancer with BRAFV600E mutation treated with BRAF/MEK-targeted therapy plus cet...	World Journal of Gastrointestinal Oncology	2025
40106946	Ileal metastasis of colorectal cancer diagnosed by double-balloon endoscopy and resected via laparoscopy: A...	International Journal of Surgery Case Reports	2025
40115908	Response to anti-EGFR therapy in chemo-refractory right-sided RAS wild-type metastatic colorectal cancer: a...	Journal of Gastrointestinal Oncology	2025
40125100	Treating a Non-mixed Neuroendocrine Colorectal Tumor Under Mixed Neuroendocrine Neoplasm (MiNEN) Protocols:...	Cureus	2025
40130050	Efficacy of disitamab vedotin-containing therapy in metastatic colorectal cancer: A case report	World Journal of Clinical Oncology	2025
40130130	Black Pigmentation on the Tongue Induced by Long-Term Use of Tetracycline Antibiotics in a Colorectal Cance...	Cureus	2025
40151739	Robotic Resection of Colonic Medullary Carcinoma: A Case Report	Cureus	2025
40161376	A case report of HER2-positive descending colon cancer with peritoneal metastasis and literature review	Frontiers in Oncology	2025
40166586	Pulmonary metastasis originating from colonic mucinous adenocarcinoma with a tree-in-bud pattern on compute...	Respiratory Medicine Case Reports	2025
40196210	Long-Term Survival after Curative Resection for Postoperative Dissemination of Pancreatic Ductal Adenocarci...	Surgical Case Reports	2025
40196211	En bloc Right Hemicolectomy with Pancreaticoduodenectomy for Advanced Ascending Colon Cancer	Surgical Case Reports	2025
40199600	Metachronous gastro-oesophageal adenocarcinoma in a patient with colon adenocarcinoma	BMJ Case Reports	2025
40206740	Colorectal cancer metastasis to the brainstem: A single case report	Surgical Neurology International	2025
40213213	Medullary carcinoma of the left colon in a 16-year-old male: an extremely rare case report from Syria	Annals of Medicine and Surgery	2025
40224173	TAS-102 in combination with bevacizumab for second-line treatment of metastatic colorectal cancer with a hy...	Frontiers in Oncology	2025
40225816	Synchronous Primary Malignancies: Incidental Detection of Ascending Colon Adenocarcinoma During Staging of ...	Case Reports in Oncological Medicine	2025
40235909	Precision treatment for human epidermal growth factor receptor 2-amplified advanced rectal cancer: A case r...	World Journal of Gastrointestinal Oncology	2025
40236342	A Rare Case of Spermatic Cord Metastasis Following Surgery for Transverse Colon Cancer	Cureus	2025

PMID	Title	Journal	Year
40270608	Case Report: A long-term response of immunotherapy combined with anti-angiogenesis therapy in a patient wit...	Frontiers in Oncology	2025
40275331	Advance imaging with magnetic resonance neurography for the diagnosis of unusual extensive pelvic perineura...	Journal of Medical Case Re- ports	2025
40276454	Intradural Cauda Equina Metastasis of Colorectal Ade- nocarcinoma: A Case Report	Cureus	2025
40308628	Rectal stenosis after immunotherapy in a mismatch re- pair deficient rectal cancer. Case report and review of...	Clinical and Translational Radiation Oncology	2025
40308701	Successful Treatment of Locally Advanced Microsatellite Instability-High Ascending Colon Cancer Using an Im...	Surgical Case Reports	2025
40322391	Torsion of Metastatic Ovarian Tumor Originating From Sigmoid Colon Cancer	Cureus	2025
40337545	Recurrent Tumor in Colorectal Cancer Requiring Com- bined Resection of Iliac or Femoral Vessels: Report of Fo...	Surgical Case Reports	2025
40342570	Rare Case of Additional Ileocecal Resection for Ascend- ing Colon Cancer with R1 Resection due to Advanced Pe...	Surgical Case Reports	2025
40351380	Partial response to posterior line immunotherapy for more than 15 months in a pMMR patient with cutaneous m...	Therapeutic Advances in Gas- troenterology	2025
40406456	Ileal pouch-anal anastomosis in familial adenomatous polyposis and ulcerative colitis with coexisting color...	Journal of Surgical Case Re- ports	2025
40416862	Immunotherapy combined with chemotherapy: a promis- ing therapeutic approach in the management of colonic squ...	Frontiers in Oncology	2025
40448204	Soft tissue metastasis in colorectal cancer: a case report	Journal of Medical Case Re- ports	2025
40452711	Robot-Assisted Multiquadrant Operation: Proctocolec- tomy and Abdominoperineal Resection Using the Da Vinci X...	Cureus	2025
40482363	A new way to remedy anastomotic leakage after TaTME: A case report	International Journal of Surgery Case Reports	2025
40491286	Histomorphological and Molecular Features of Colonic Large-Cell Neuroendocrine Carcinoma in a Patient With ...	International Journal of Sur- gical Pathology	2025
40519284	Simultaneous double primary malignant tumors of MSS/pMMR ascending colon cancer and MSI-H/dMMR duodenal can...	Frontiers in Oncology	2025
40519480	Acute Perforated Large Bowel Obstruction Due to Col- orectal Malignancy: A Surgical Emergency With Options	Cureus	2025
40524862	Immune Checkpoint Inhibitors for Microsatellite Insta- bility High Unresectable Obstructive Colon Cancer: A R...	Surgical Case Reports	2025
40526090	Case Report WIN-MTB-2023001 WIN International Molecular Tumor Board A 62-year-old male with metastatic colo...	Oncotarget	2025
40532548	A rare case of synchronous anal metastasis from sigmoid adenocarcinoma: case report and literature review	International Journal of Surgery Case Reports	2025
40546735	Successful sphincter preservation and long-term survival in a patient with ultra-low locally advanced recta...	Therapeutic Advances in Medical Oncology	2025
40547162	Primary squamous cell carcinoma of the descending colon with pancreatic metastasis: A case report	World Journal of Gastroin- testinal Oncology	2025
40552007	Two Cases of Small Intestinal Follicular Lymphoma Pre- senting with Intestinal Stricture	Surgical Case Reports	2025
40556307	Colorectal Cancer with Ovarian Metastasis After Pani- tutumab and FOLFOX4: A Case Report	The American Journal of Case Reports	2025
40568773	Multimodal therapy and precision-imaging extends the limits of treatment for a patient with initially unres...	Annals of The Royal College of Surgeons of...	2025
40578238	Solitary metastases of colon adenocarcinoma to the an- kle: A case report and literature review	International Journal of Surgery Case Reports	2025

PMID	Title	Journal	Year
40589557	Neoadjuvant Chemotherapy for Stage II Rectal Cancer Diagnosed in the Second Trimester of Pregnancy: A Case ...	Surgical Case Reports	2025
40599858	Diffuse large B-cell lymphoma combined with two solid tumors: a case report	Frontiers in Oncology	2025
40603091	Four Metachronous Cancers: Mutation Profiling Can Be Used to Distinguish Multiple Primary Cancers from Meta...	Internal Medicine	2025
40606991	Case Report: Metastatic colorectal cancer with ALK-CEP44 fusion and rapid resistance development	Frontiers in Oncology	2025
40613034	Radical treatment for metastasis of HER2-positive rectal adenocarcinoma to the liver: A case report and lit...	Oncology Letters	2025
40625540	Long-Term Survival after Choroidal Metastasis Following Colorectal Cancer Surgery: A Case Report	Surgical Case Reports	2025
40677671	Robotic-Assisted Right Upper Lobe Pulmonary Metastectomy in a Patient with Rectal Adenocarcinoma: a Case ...	Mædica	2025
40688855	Successful Treatment of BRAF V600E-Mutant Intrahepatic Cholangiocarcinoma With Encorafenib, Binimetinib, an...	Cureus	2025
40696657	Appendiceal goblet cell adenocarcinoma synchronous with ascending colon adenocarcinoma and severe anemia: A...	Medicine	2025
40709126	Silent Yet Striking: Medullary Carcinoma Behind an Intestinal Obstruction	Cureus	2025
40713591	Cervical and vaginal metastases from ascending colon cancer: a case report and literature review	BMC Women's Health	2025
40735333	Harnessing immunotherapy in sporadic MSI-H/dMMR colorectal cancer: a case study	Journal of Surgical Case Reports	2025
40740271	Rare Case of Duodenal Metastasis From Colon Cancer: Review of Literature and Insights on Novel Therapies	Case Reports in Gastrointestinal Medicine	2025
40753237	Synchronous triple-negative breast cancer, rectal adenocarcinoma, and chemotherapy-induced hepatitis B virus...	Journal of Medical Case Reports	2025
40761192	Perforated primary adenocarcinoma of the colon with choriocarcinoma differentiation treated with targeted c...	AME Case Reports	2025
40761477	Surgical Resection of a Solitary Metachronous Retrocaval Lymph Node Metastasis above the Renal Vein from Re...	Surgical Case Reports	2025
40777125	Case Report: A case of interstitial pneumonia in a colorectal cancer patient with a history of lung cancer ...	Frontiers in Oncology	2025
40777693	Non-operative Management of Locally Advanced Rectal Adenocarcinoma Using a Watch-and-Wait Approach: A Repor...	Cureus	2025
40786361	Twists of the Gut and Genome: A Case of Intestinal Intussusception Revealing Lynch Syndrome in a Young Adul...	Cureus	2025
40797474	Robot-assisted surgery for rectal cancer with tumor-associated dermatomyositis during preoperative chemothe...	Medicine	2025
40814385	HER2-positive, RAS-mutant, MSS colorectal cancer: a rare subtype report and novel insights into immunothera...	American Journal of Cancer Research	2025
40826135	Postpartum colonic squamous cell carcinoma with a BRAFV600E mutation: a case report	Journal of Medical Case Reports	2025
40837582	Case Report: Successful treatment of triple primary cancers (lung, stomach, and intestine) using sintilimab...	Frontiers in Immunology	2025
40842405	Immunotherapy With Pembrolizumab for Resectable dMMR/MSI-H Stage III Colon Cancer: A Case of Personalized, ...	Journal of Investigative Medicine High Imp...	2025
40863225	An Uncharted Path of Metastasis: A Case Report of Sigmoid Colon Cancer with Synchronous Vaginal and Urethra...	Diseases	2025
40878818	Multi-omic characterization of consensus molecular subtype 1 (CMS1) colorectal cancer with dampened immune ...	Molecular Oncology	2025

PMID	Title	Journal	Year
40909950	Combined neoadjuvant chemotherapy and immunotherapy in a hepatitis B virus-positive patient with locally ad...	Frontiers in Oncology	2025
40917781	Solitary pulmonary metastasis with cystic airspaces in colon cancer: A rare case report	Open Life Sciences	2025
40935920	Medullary Carcinoma of the Colon: Not as Favourable as Previously Thought?	Journal of Gastrointestinal Cancer	2025
40951242	Emergency Colectomy for Obstruction in Microsatellite Instability Colonic Cancer: A Complete Response Follo...	Cureus	2025
40963870	An innovative treatment approach using iliac artery embolization and implanted port for chemotherapy in a p...	Frontiers in Oncology	2025
40995568	Renal Metastasis of Adenocarcinoma of the Colon	Case Reports in Pathology	2025
41010809	Robotic-Assisted Total Pelvic Exenteration for Rectal Cancer Using the Hugo™ RAS System: First Case Report	Journal of Clinical Medicine	2025
41019039	Long-term survival benefit from third-line treatment for advanced colon cancer: a case report	Frontiers in Immunology	2025
41024295	Advanced-stage colorectal cancer in the context of coronavirus disease-19, highlighting socioeconomic dispa...	Journal of Medical Case Reports	2025
41030488	A Forty-Three-Year-Old Male With Penile and Cavernous Metastases From Rectal Cancer	Journal of Medical Cases	2025
41030825	Four Surgical Cases of Medullary Carcinoma of the Colon	Surgical Case Reports	2025
41043274	Metachronous breast and rectal cancer: A clinical case report and review	International Journal of Surgery Case Reports	2025
41054660	Successful Treatment of Disseminated Carcinomatosis of the Bone Marrow Secondary to Colorectal Cancer With ...	Cureus	2025
41077740	Diagnostic Challenges in Sessile Serrated Lesions: Progression to Adenocarcinoma in a High-Risk Patient	The American Journal of Case Reports	2025
41094574	Rare double primary carcinoma in ileocecal region: cecal tubular adenocarcinoma combined with appendiceal g...	Diagnostic Pathology	2025
41107652	Role of surgical resection for non-colorectal non-neuroendocrine liver metastases	World Journal of Hepatology	2017
41111754	Late-Onset Isolated Drain Site Metastasis in Colorectal Cancer Managed by a Multidisciplinary Approach	Cureus	2025
41113075	Spontaneous colonic transection following pathologic complete response to pembrolizumab in high microsatell...	World Journal of Clinical Cases	2025
41127008	Case Report: Successful treatment of a case of Lynch syndrome with double primary ovarian and rectal cancer	Frontiers in Oncology	2025
41127013	Case Report: Failed response to anti-PD-1 immunotherapy in a colon cancer patient with high microsatellite ...	Frontiers in Oncology	2025
41127750	Pancreas-Sparing Duodenal Resection in Colorectal Adenocarcinoma With Local Invasion of the Duodenum: A Cas...	Cureus	2025
41164049	Mandibular Metastasis From Colon Cancer Presenting as Jaw Pain	Cureus	2025
41169286	Rare but overlooked: choroidal metastasis as a late manifestation of colon cancer with review of literature	Oxford Medical Case Reports	2025
41170468	Case Report: Unique immunotherapy response in a patient with metachronous colorectal cancer not associated ...	Frontiers in Oncology	2025
41190262	Anatomical Hepatectomy in a Patient with Situs Inversus Totalis and Right-Sided Round Ligament	Surgical Case Reports	2025
41245504	Noninvasive Therapeutic Monitoring of Circulating Tumor DNA in BRAF-Mutant Metastatic Colon Cancer Using Dr...	Case Reports in Oncological Medicine	2025
41258467	Colorectal adenocarcinoma in children and adolescents: the management of advanced disease	International Journal of Colorectal Disease	2025
41310725	Aberrant TTF-1 expression in metastatic colorectal adenocarcinoma mimicking primary lung cancer: a case rep...	Diagnostic Pathology	2025

PMID	Title	Journal	Year
41325200	Neoadjuvant immunotherapy for dMMR/MSI-H locally advanced rectal cancer patients: demystifying the 100% cl...	The Oncologist	2025
41332047	Case report of sigmoid colon cancer with synchronous peritoneal metastasis achieving a pathological complet...	International Journal of Surgery Case Reports	2025
41332966	Atypical bone and synovial metastases from colorectal adenocarcinoma: A case report	Radiology Case Reports	2025
41346621	Rare MSI-H hepatoid adenocarcinoma of the colon with BRAF V600E mutation achieving long-term disease-free s...	Frontiers in Immunology	2025
41355881	Squamous cell carcinoma of the sigmoid colon: a rare case report	Journal of Surgical Case Reports	2025
41355935	Unlocking New Treatment Possibilities for Metastatic Endometrial Cancer With KRAS G12C Mutation	Case Reports in Oncological Medicine	2025
41368535	A New Hope: Dual Human Epidermal Growth Factor Receptor 2 Blockade and Chemotherapy in Human Epi-dermal Grow...	Case Reports in Oncology	2025
41378290	A case report of multiple primary malignant tumors: breast cancer and colorectal cancer	Frontiers in Oncology	2025
41423794	Metastatic colorectal adenocarcinoma with APC and EGFR L858R co-mutations: a case report	The Oncologist	2025
41425970	Peritoneal metastases from colorectal cancer managed by stereotactic body radiation therapy: presentation o...	BJR Case Reports	2025
41439088	Efficacy of Pembrolizumab as Conversion Therapy in De-ficient Mismatch Repair/Microsatellite Instability-Hig...	Cureus	2025
41458342	A Pathological Complete Response After Pembrolizumab for DNA Mismatch Repair-Deficient/Microsatellite In-sta...	Cureus	2025
41458856	An Older Woman With Postoperative Breast Metastasis From Ascending Colon Cancer Successfully Treated With P...	Cureus	2025
41466910	Bevacizumab-Associated Bowel Perforation Causing Evisceration	Cureus	2025
41487331	Delayed Presentation of Metachronous Ascending Colonic Adenocarcinoma in Early-Onset Colorectal Can-cer Unde...	Journal of Medical Cases	2025
41487569	Sensitivity of metastatic mucinous tailgut cyst adenocar-cinoma to gemcitabine and radiation: a case report	Frontiers in Oncology	2025
41492327	Isolated Splenic Metastasis With Perisplenic Artery Lymph Node Involvement More Than 10 Years After Cu-rativ...	Cureus	2026
41497708	Protein-Losing Enteropathy Caused by Ascending Colon Cancer without Polyposis: A Rare Case Diagnosed by Sci...	Surgical Case Reports	2025
41503128	Neoantigen dendritic cell vaccination combined with con-ventional chemotherapy for a patient with advanced t...	Journal of Surgical Case Re-ports	2026
41522095	Synchronous Retroperitoneal Pleomorphic Liposarcoma and Colonic Mucinous Adenocarcinoma	Case Reports in Oncology	2025
41522807	Curative surgery enabled by radiotherapy in rectal can-cer: A case report	Oncology Letters	2025
41541147	Case report: Secondary peristomal cutaneous squamous cell carcinoma following multimodal treatment for rect...	International Journal of Surgery Case Reports	2025
41555436	Rectal melanoma: managing a rare cause of rectal bleed-ing; case report and review of the literature	World Journal of Surgical On-cology	2026
41561736	Case report: Minimally invasive management of syn-chronous early-stage ascending colon adenocarcinoma and ty...	Frontiers in Oncology	2026
41561746	Neoadjuvant therapy with cadonilimab in a patient with MSI-H/dMMR colorectal cancer: a case report	Frontiers in Oncology	2026
41561768	Case Report: Pulmonary enteric adenocarcinoma har-boring KRAS/TP53/APC mutations with contralateral lung met...	Frontiers in Oncology	2026
41573493	Perianal Paget's Disease From Underlying Anal Adeno-carcinoma: A Case Study	Cureus	2025

PMID	Title	Journal	Year
41602383	Case Report: Elevated myocardial enzymes in a patient with metastatic rectal cancer without myocardial injury	Frontiers in Oncology	2026
41613130	Lung adenocarcinoma and colorectal cancer as double primary malignancies reveal lynch syndrome: a case repo...	Frontiers in Immunology	2026
41623610	Concomitant RAS and BRAF V600E Mutations in Unresectable Metastatic CRC: A Case of Long-term Survival	Journal of the Anus, Rectum and Colon	2026
41647788	Robotic-Assisted Sigmoidectomy with D3 Lymphadenectomy for Sigmoid Colon Cancer in a Patient with Situs Inv...	Surgical Case Reports	2026
41658037	Pathological Complete Response Achieved with Colorectal Cancer-Based Chemotherapy for Locally Recurrent Cec...	Surgical Case Reports	2026
41684994	Adult Pan-Colonic Intussusception Secondary to Poorly Differentiated Cecal Adenocarcinoma: A Rare Presentat...	Cureus	2026
41695926	Solitary esophageal metastasis ten years after curative resection of stage I rectal adenocarcinoma: A case ...	World Journal of Gastrointestinal Oncology	2026

Supplementary S5: Custom Drug-Class Mapping

The deterministic scorer’s class-level matching for CRC therapy regimens uses a curated drug-to-class dictionary aligned with standard oncology pharmacology nomenclature. The complete mapping is reproduced below; drug names within each class are normalized to lower case and matched after alias resolution (e.g., FOLFOX4 → folfox, Lonsurf → ftd/tpi). The source-of-truth implementation is in the released code repository.

Drug class (scorer tag)	Member drugs / aliases
ALK_inhibitor	alectinib, crizotinib, ensartinib, lorlatinib
BCL2_inhibitor	venetoclax
BRAF_inhibitor	brafotvi, dabrafenib, encorafenib, tafinlar, vemurafenib, zelboraf
CD20_mAb	rituximab
CD30_ADC	bv-avd
EGFR_TKI_1G	gefitinib
EGFR_TKI_2G	afatinib
EGFR_TKI_3G	osimertinib
GCSF	filgrastim, g csf, g-csf, gm csf, gm-csf
KIT_inhibitor	imatinib
KRAS_G12C_inhibitor	adagrasib
MEK_inhibitor	binimetinib, trametinib
MET_inhibitor	crizotinib
NTRK_inhibitor	larotrectinib
PARP_inhibitor	olaparib, talazoparib
RANK_L_mAb	denosumab
RET_inhibitor	pralsetinib, selpercatinib
ROS1_inhibitor	crizotinib, lorlatinib
VEGFR_TKI_multikinase	anlotinib, apatinib, cabozantinib, fruquintinib, lenvatinib, regorafenib, so- rafenib
anti_CTLA4	anti ctla4, anti-ctla-4, ctla-4 inhibitor, ici, iparomlimab, ipilimumab
anti_EGFR_mAb	anti egfr, anti-egfr, cetuximab, egfr antibody, panitumumab
anti_HER2_ADC	disitamab vedotin, t-dxd
anti_HER2_TKI	pyrotinib
anti_HER2_mAb	pertuzumab, trastuzumab
anti_PD1	anti pd-1, anti-pd-1, anti-pd-1 immunotherapy, camrelizumab, dostarlimab, ici, immune checkpoint inhibitor, immune checkpoint inhibitors, nivolumab, pd-1 inhibitor, pd-1 monoclonal antibody, pd1 inhibitor, pembrolizumab, ser- plulimab, sintilimab, tislelizumab, toripalimab, tuvonralimab
anti_PD1_CTLA4_bispecific	cadonilimab
anti_PDL1	anti pd-l1, anti-pd-l1, atezolizumab, envafolimab, ici, immune checkpoint in- hibitor, immune checkpoint inhibitors, pd-1/pd-l1 inhibitor, pd-l1 inhibitor, pd1 pdl1 inhibitor
anti_VEGF_fusion	afibercept
anti_VEGF_mAb	anti vegf, anti-vegf, bevacizumab, ramucirumab, vegf antibody
bisphosphonate	zoledronic acid
chemo_alkylating	temozolomide
chemo_anthracycline	doxorubicin
chemo_antimetabolite	5 fluorouracil, 5 fu, 5-fu, capecitabine, cytarabine, fluoropyrimidine, fluo- rouracil, gemcitabine, gimeracil, ldac, oteracil, raltitrexed, s-1, single-agent fluoropyrimidine therapy, tas-102, tegafur, tipiracil, trifluridine, trifluridine tipiracil, trifluridine-tipiracil, trifluridine/tipiracil
chemo_combination_regimen	abvd, avd, bv-avd, capeox, capox, cmfolfoxiri, folfiri, folfox, folfox6, folfoxiri, ftd/tpi, mfolfox, mfolfox-6, mfolfox6, platinum-based combination chemother- apy, r-chop, r-dhap, xelox
chemo_other	leucovorin, mitomycin
chemo_platinum	carboplatin, cisplatin, oxaliplatin, platinum-based combination chemotherapy
chemo_taxane	abraxane, docetaxel, nab-paclitaxel, paclitaxel
chemo_topoisomerase1	irinotecan
chemo_topoisomerase2	etoposide
hormone_CYP17_inhibitor	abiraterone
hypomethylating_agent	hma
proteasome_inhibitor	bortezomib
supportive_care_other	calcitonin, fmt, imipenem, octreotide, sulfamethoxazole, trimethoprim

Supplementary S6: Reproducibility

This appendix documents the cryptographic lock and end-to-end reproduction protocol that supports the bit-for-bit reproducibility claim in Methods §2.3.

S6.1 SHA-256 hash-locking

Every artifact whose content could influence a scored verdict is bound to a SHA-256 hash recorded in the release manifest. The hashed artifacts are:

- **Benchmark items:** each of the 1,586 NCCN structured items and 419 CRC case items, individually hashed and aggregated into a benchmark hash file released with the paper.
- **Gold standard:** the full gold-standard file (one hash for the file, plus per-item canonical-form hashes).
- **Source manifests:** the NCCN workspace registries and CRC case registries that drive item generation.
- **Workflow prompts:** every prompt template used in the generation pipeline, including system prompts and per-step prompts.
- **Scorer source:** the deterministic scorer source, its JSON schemas, and the drug-class dictionary.
- **Release manifest:** a master manifest that records all of the above hashes plus the release tag.

A scorer run is "verified" only when (i) the release manifest hash matches; (ii) the input hashes in the manifest match the on-disk inputs; (iii) the scorer source hash matches the released scorer. Any mismatch flags the run as non-canonical and is reported in the scorer's stdout log.

S6.2 Bit-for-bit reproducibility guarantee

Given a fixed manifest hash, the scorer is bit-for-bit reproducible: running the released scorer on the same set of raw model-output records produces byte-identical item-score and summary files across operating systems and Python patch versions, modulo JSON whitespace normalization (which the scorer pins to a fixed representation). This property is enforced by three mechanisms:

- No randomness in the scorer code path (no **random**, no temperature, no sampling).
- No external API calls during scoring (no LLM-as-judge, no network IO).
- Canonicalization of allowed schema repairs to a fixed normal form before label assignment.

S6.3 Computational environment

The reported results were produced under the following environment.

Component	Version
Python interpreter	3.11
NumPy	1.26
SciPy	1.11
jsonschema	≥ 4.0
matplotlib	≥ 3.7 (for figure scripts)
Operating system	Ubuntu 22.04 LTS

A pinned **requirements.txt** is released at the canonical archive. The scorer itself depends only on the Python standard library plus **jsonschema**; the broader analysis (bootstrap, permutation, figure generation) depends on NumPy, SciPy, and matplotlib.

S6.4 End-to-end reproduction protocol

The following protocol regenerates every figure and table from the released artifacts. We summarize each step; full commands and expected outputs are documented in the public release README.

1. **Clone the public release:** `git clone https://github.com/dcszhang/Benchmark` and check out the release tag (e.g., `v1.0.0`); the canonical commit hash is recorded in the release manifest and on the Zenodo archive page.

2. **Install the pinned environment:** `pip install -r requirements.txt` in a fresh Python 3.11 environment.
3. **Verify integrity:** a release-verification script confirms that the on-disk hashes match the release manifest.
4. **Re-score model outputs:** running the released scorer on the raw model outputs reproduces item-level scores, the summary file, and the failure-label distribution per model.
5. **Re-aggregate statistics:** the analysis scripts in `analysis/` consume the per-model scores and reproduce the bootstrap CIs, permutation p -values, and adjacency-tier assignment.
6. **Re-render figures and tables:** the figure scripts in `analysis/` consume the aggregated statistics and emit the figures and tables in the manuscript.

S6.5 Paper-of-record identifier

The canonical commit hash and Zenodo DOI corresponding to the version of the artifacts used in this manuscript are recorded in the manuscript’s Data Availability and Code Availability statements. Future updates to the public release will be tagged with new identifiers; the paper-of-record identifier is immutable.

S6.6 Model registry: OpenRouter identifiers

The nine evaluated models were accessed through the OpenRouter API using the exact identifiers below. The same registry is encoded in the released code.

Model (display name in the main paper’s Models evaluated table)	OpenRouter identifier
Claude Sonnet 4.6	<code>anthropic/claude-sonnet-4.6</code>
GPT-5.5	<code>openai/gpt-5.5</code>
Gemini 2.5 Pro	<code>google/gemini-2.5-pro</code>
Gemini 3.1 Pro Preview	<code>google/gemini-3.1-pro-preview</code>
Qwen 3.6-Plus	<code>qwen/qwen3.6-plus</code>
DeepSeek V4 Pro	<code>deepseek/deepseek-v4-pro</code>
GLM-5	<code>z-ai/glm-5</code>
GLM-5.1	<code>z-ai/glm-5.1</code>
Minimax M2.7	<code>minimax/minimax-m2.7</code>

Each identifier resolves to a specific provider checkpoint at evaluation time. Independent re-querying of the same checkpoint may yield different behavior because of provider-side version drift between vendor updates.

Supplementary S7: Gold-Answer Audit on the Zero-Correct Subset

This appendix documents the second-wave clinician audit referenced in the Methods (Clinician adjudication) and Results (Clinician validation) of the main manuscript. The audit’s purpose is distinct from the first-wave adjudication (Supplementary S1): rather than evaluating the deterministic scorer’s verdict given a fixed gold standard, this audit asks two independent oncologists to directly evaluate the *gold standard itself* on the items that underpin the 42% capability boundary.

S7.1 Sampling design

From the 845 items in the zero-correct set, we drew a stratified random sample of $n = 80$ items using a frozen seed (`seed = 42`). The stratum allocation is shown below.

Stratum	n	% of zero-correct parent
Universal-fail (all 9 models incorrect)	9	1.1%
In-guide handoff resolution	20	25.0%
Upstream routing decision	15	18.8%
Missing information request	15	18.8%
Parallel option disambiguation	15	18.8%
Evidence resolution request	3	3.8%
Colorectal cancer (CRC) cases	3	3.8%
Total	80	100%

Design intent: (i) include *all* nine universal-fail items (the hardest subset underpinning the 42% boundary, which should be fully covered by gold-answer audit); (ii) oversample question types most likely to be challenged by reviewers (in-guide handoff, upstream routing) to maximize the chance of surfacing weak gold standards; (iii) retain small counts of evidence-resolution and CRC strata for existence-level signal on edge tracks.

S7.2 Audit protocol

Two board-certified oncologists, independent of the first-wave adjudication (i.e., different individuals) and blinded from each other’s responses, each rated every item on two orthogonal dimensions:

1. **Clinical correctness:** is the NCCN gold answer clinically correct? Values: `correct` / `arguable` / `wrong`.
2. **NCCN traceability:** can the gold answer be located in the corresponding workspace’s NCCN guideline at the target pathway node? Values: `traceable` / `partially` / `not_traceable`.

Reviewers were *not* shown any model output to avoid contamination. The auditors had access to the official NCCN guideline PDF for each item’s cancer workspace.

S7.3 Headline results

Metric	Value	Wilson 95% CI
Both reviewers labeled gold <code>wrong</code>	2 / 80 (2.5%)	[0.7%, 8.7%]
Either reviewer labeled gold <code>wrong</code>	3 / 80 (3.8%)	[1.3%, 10.5%]
Both reviewers labeled gold <code>traceable</code>	76 / 80 (95.0%)	[87.8%, 98.0%]
<code>wrong</code> flags on all 9 universal-fail items	0 / 9	—

Key observations: (i) the gold-error rate on the zero-correct subset is bounded above by approximately 4%–11% even under the most permissive definition, well below the threshold at which the 42% wall could be re-explained as a gold-quality artifact; (ii) no universal-fail item received a `wrong` flag from either reviewer—all 9 were unanimously labeled `correct` or `arguable`, indicating that the hardest subset of the 42% wall is precisely where the gold standard is most defensible.

S7.4 Inter-rater agreement

Dimension	Weighted κ	95% bootstrap CI
NCCN traceability	0.740	[0.25, 1.00]
Clinical correctness	0.353	[0.13, 0.54]

Traceability $\kappa = 0.740$ falls in the substantial-agreement range. Clinical correctness $\kappa = 0.353$ falls in the fair-agreement range, but requires careful interpretation:

The lower clinical-correctness κ is driven almost entirely by a single-direction systematic shift: 23 of 27 disagreements (85%) take the form “Reviewer A: **arguable**, Reviewer B: **correct**”, reflecting differing thresholds for the intermediate **arguable** category rather than substantive clinical disagreement. **No item was labeled correct by one reviewer and wrong by the other.** That is, no bidirectional extreme disagreement was observed.

Practical interpretation: the two reviewers’ disagreements primarily reflect “is this gold answer uncontestedly correct, or does it fall in a class of cases that could be reasonably questioned”—not “is this gold answer right or wrong.” Even the stricter reviewer did not upgrade any gold to **wrong**.

S7.5 Per-stratum breakdown

Stratum	n	Both wrong	Either wrong
Universal-fail (all 9 models)	9	0	0
In-guide handoff resolution	20	0	1 (5.0%)
Upstream routing decision	15	0	0
Missing information request	15	0	0
Parallel option disambiguation	15	0	0
Evidence resolution request	3	0	0
Colorectal cancer (CRC) cases	3	2 (66.7%)	2 (66.7%)
Total	80	2 (2.5%)	3 (3.8%)

Two observations are worth highlighting:

(1) NCCN-anchored items (first 6 strata, $n = 77$) have very high gold quality: across the entire NCCN subset of 77 items, only 1 was labeled **wrong** by either reviewer (the in-guide handoff stratum, 1/20). This is consistent with intuition: NCCN items are derived from formally published guidelines, with gold answers anchored directly to specific guideline nodes; the room for “gold error” is correspondingly small.

(2) The CRC case stratum ($n = 3$) shows substantially lower gold quality: 2 of 3 CRC items were labeled **wrong** or not-traceable by both reviewers. This is consistent with the character of the CRC track: CRC gold answers are derived from actual published case-report treatments rather than NCCN-recommended pathways. Published case-report treatments often include trial-stage drugs or individualized regimens that do not necessarily reflect standard NCCN recommendations.

This observation provides independent empirical support for the manuscript’s separation of NCCN and CRC findings throughout the Results: the CRC track measures whether LLMs can recommend “actually used but outside-NCCN” regimens in a real-world clinical context, rather than whether LLMs comply with NCCN guidelines.

S7.6 Data release

The full audit pack, per-reviewer item-level responses (including optional comments), the side-by-side reviewer comparison with disagreement flags, the sampling manifest with frozen seed, and the analysis scripts that produced every table in this appendix are released at the canonical archive linked from the paper’s Data Availability statement (<https://github.com/dcszhang/Benchmark>).

References

- [1] Edwin B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927.
- [2] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.