GBM Clinical-Trial Eligibility — Reconciled Confidence Table

Across 25 glioblastoma trials | Claude analysis reconciled against ChatGPT estimates

How to read this. A single confidence number hides two different things, so they are split here. **Prevalence (n/25)** = how many of the 25 trials impose the criterion at all (a prior probability for an unseen GBM trial). **Hard-gate confidence** = given a trial states it, how deterministically it screens a patient in/out (objective numeric thresholds score high; investigator-discretion language scores low). The **ChatGPT** column is the prior estimate being reconciled; the Verdict explains any divergence. Key pattern: ChatGPT's single column mixes the two axes — for non-universal criteria its number behaves like prevalence, which makes hard but population-specific gates (disease stage, biomarkers, organ-function labs) look soft.

TIER A — Near-universal inclusion gates (high prevalence + high strictness)

Eligibility criterion	Prevalence (n/25)	Hard-gate confidence (when stated)	ChatGPT score	Verdict / correction
GBM / glioma diagnosis confirmed (histologic or radiographic)	25/25	99%	98%	Agree. Definitional. Nuance: drug/cell trials demand histologic or molecular proof; imaging & surgical trials accept radiographic appearance.
Age ≥18 years	23/25	98%	98%	Agree. Variants: 19 (Trial 16, Nebraska), ≥22 (Trial 11); upper caps 80–90 in Trials 12, 13, 20, 23.
Informed consent / ability to comply	25/25	99% required	98%	Required by all, so 98% is fair — but near-zero screening value: it almost never excludes a clinically eligible patient. Do not weight it in a matching score.
Performance status (KPS or ECOG)	16/25	95%	95%	Strictness agrees, but it is not universal (~64%): imaging, observational & surgical trials (7, 8, 12, 13, 14, 24) omit it. Threshold varies: KPS ≥60/70, ECOG 0-1/0-2/0-3.

TIER B — Conditional inclusion gates (population/modality-specific: hard when in scope, partial prevalence)

Eligibility criterion	Prevalence (n/25)	Hard-gate confidence (when stated)	ChatGPT score	Verdict / correction
Adequate hematologic function (ANC, platelets, Hgb)	11/25	97%	93%	ChatGPT overstates prevalence. Numeric panels appear only in drug/cell-therapy/interventional trials; absent from imaging, surgical & observational ones.
Adequate renal function (creatinine / CrCl)	11/25	97%	92%	Same as above — prevalence ~44%, not ~92%. Hard gate when present.
Adequate hepatic function (bilirubin, AST/ALT)	11/25	97%	92%	Same as above.
Disease stage = newly diagnosed required	~7/25	99%	35%	**Axis confusion.** 35% reads as prevalence; when a trial specifies stage it is a near-absolute gate. Decisive for matching.
Disease stage = recurrent / progressive required	~10/25	99%	65%	**Axis confusion**, same as above. Stage is one of the hardest gates; it just isn't universal because trials split by population.

TIER B — Conditional inclusion gates (population/modality-specific: hard when in scope, partial prevalence)				
Eligibility criterion	**Prevalence (n/25)**	**Hard-gate confidence (when stated)**	**ChatGPT score**	**Verdict / correction**
Prior radiation + temozolomide completed	~9/25	95%	50%	Hard, objective gate in recurrent-disease trials (washout windows). ChatGPT undersells strictness.
Measurable disease on MRI (RANO)	~6/25	92%	55%	Lower prevalence than implied; objective and strict when stated (size/volume bounds).
Tumor size / volume limit	~6/25	95%	40%	Objective hard gate where present (e.g. ≤3.5 cm T1, <6 cm T6, 2–20 cm3 T20); uncommon overall.
Surgical candidacy / resection planned	~11/25	90%	45%	Prevalence ~44%; a firm gate when in scope (surgical, device, tissue-collection trials).
Stable / decreasing steroid dose (≤6–8 mg dex)	9/25	90%	75%	ChatGPT overstates prevalence and understates strictness; objective and mechanistically enforced where stated.
Recovery from prior-therapy toxicity (≤Grade 1)	7/25	85%	68%	Prevalence lower than implied; CTCAE-graded, fairly objective with standard alopecia/neuropathy carve-outs.
Negative pregnancy test (if childbearing potential)	18/25	95%	90%	Agree. Objective test.
Contraception agreement (if childbearing potential)	18/25	80%	88%	Slightly softer than ChatGPT — largely an attestation rather than a verifiable gate.
Life expectancy ≥12 wk / ≥3 mo	4/25	55%	80%	**Overstated on both axes.** Only ~4 trials state it, and it is a subjective clinician estimate, not a precise filter.
Seizures controlled if present	~4/25	80%	25%	Reasonably hard when stated (stable AED regimen); uncommon. ChatGPT too low on strictness.

TIER C — Common exclusions				
Eligibility criterion	**Prevalence (n/25)**	**Hard-gate confidence (when stated)**	**ChatGPT score**	**Verdict / correction**
Pregnancy / lactation (exclusion)	18/25	95%	95%	Agree.
Other / active secondary malignancy	14/25	80%	40%	ChatGPT understates prevalence (~56%). Consistent carve-outs (in-situ, basal/squamous skin, disease-free ≥3 yr) soften it modestly.
HIV / immunodeficiency	9/25	90%	35%	Prevalence agrees (~36%); objective serology, hard gate, concentrated in cell-therapy / immunotherapy / myelosuppressive trials (3, 9, 17, 19).
Active hepatitis B / C	9/25	90%	(folded into infection)	Treat separately from generic infection: objective serology, hard gate. Same trial cluster as HIV.

TIER C — Common exclusions				
Eligibility criterion	**Prevalence (n/25)**	**Hard-gate confidence (when stated)**	**ChatGPT score**	**Verdict / correction**
Active infection requiring systemic treatment	~10/25	70%	70% / 45%	Agree-ish; partly qualitative ('requiring systemic Rx').
MRI / gadolinium contraindication, pacemaker	13/25	95%	70% / 65%	Strict, objective safety gate; prevalence ~52% and dominant in imaging/device trials (7, 8, 11, 20, 21).
Implanted device / skull defect / metallic objects	8/25	92%	(not scored)	Objective; specific to device, ultrasound, TTFields & MRI trials (8, 11, 20). Worth adding to the model.
Tumor location: multifocal / leptomeningeal / infratentorial / midline / 'butterfly'	12/25	90%	(listed for algo)	Objective radiographic call; trial- and modality-specific. Belongs in the high-weight set, as ChatGPT's algorithm list notes.
Recent conflicting investigational therapy / washout	~12/25	88%	60%	Washout windows are objective and hard; ChatGPT understates strictness.
Prior anti-PD-1 / PD-L1 / checkpoint therapy	3/25	95%	(not scored)	Objective hard gate, but only in immunotherapy trials (1, 5, 6).
Significant cardiovascular disease	10/25	70%	(not scored)	Mixed: objective where time-bound (MI/CVA <6 mo), qualitative where 'significant/uncontrolled'.
Autoimmune disease / immunosuppression / chronic high-dose steroids	5/25	80%	(not scored)	Objective-ish with explicit carve-outs; almost entirely confined to checkpoint-inhibitor trials.

TIER D — Soft / discretionary clauses (stated often, low determinism — do not weight heavily)				
Eligibility criterion	**Prevalence (n/25)**	**Hard-gate confidence (when stated)**	**ChatGPT score**	**Verdict / correction**
Severe / uncontrolled intercurrent medical illness	12/25	45%	85%	**Overstated.** This is the discretionary 'in the opinion of the investigator' catch-all — stated often, but it is judgment, not a deterministic gate.
Blanket investigator-discretion clause ('any condition that, in the investigator's opinion...')	≥8/25	25%	(not scored)	Genuinely present (the only exclusion in Trial 14) but unscreenable. Effectively zero algorithmic value.
Psychiatric illness / social situation limiting compliance	~8/25	40%	(not scored)	Discretionary; low determinism.

TIER D — Soft / discretionary clauses (stated often, low determinism — do not weight heavily)				
Eligibility criterion	**Prevalence (n/25)**	**Hard-gate confidence (when stated)**	**ChatGPT score**	**Verdict / correction**
Inadequate organ function (as exclusion)	11/25	90%	90%	Agree on strictness, but prevalence-limited — mirrors the lab-panel trials only.
Poor performance status (as exclusion)	16/25	90%	90%	Mirror of the inclusion PS gate; same ~64% prevalence.

TIER E — Molecular / biomarker requirements (low prevalence, near-absolute when present)				
Eligibility criterion	**Prevalence (n/25)**	**Hard-gate confidence (when stated)**	**ChatGPT score**	**Verdict / correction**
MGMT promoter status requirement	~1/25	99%	15%	As a true requirement rarer than 15% (Trial 17 demands unmethylated). Hard gate when present; elsewhere MGMT is recorded, not required.
IDH mutation / wild-type requirement	~3/25	99%	15%	Direction differs by trial: T10 needs IDH1-mutant, T19 needs IDH-wildtype, T17 needs IDH-absent. Hard gate, low prevalence.
CD70 positivity	1/25	99%	5%	Agree. Rare (Trial 9), absolute when present.
GPC3 positivity	1/25	99%	5%	Agree. Rare (Trial 18).
B7-H3 expression	1/25	99%	5%	Agree. Rare (Trial 3).
Tumor Treating Fields willingness	1/25	90%	5%	Agree. Single trial (11).
Any specific molecular biomarker (aggregate)	~6/25	95%	20%	Low prevalence overall, but whenever a biomarker is named it is a hard, non-negotiable gate.

Recommended matching weights (two-axis view). Highest weight to gates that are both objective and decisive even if not universal: GBM diagnosis, age, performance status, organ-function labs, pregnancy status, disease stage (newly diagnosed vs recurrent), prior therapy / washout, tumor location & size, MRI eligibility, and any named biomarker. Down-weight or exclude the discretionary clauses in Tier D and treat informed consent as a pass-through. ChatGPT's algorithm shortlist is sound; the main fix is to score disease stage and biomarkers as hard gates conditioned on the trial, not as low-probability requirements.

Prevalence counts are derived from the 25 eligibility sections in the source file; '~' marks counts that depend on how borderline/either-stage trials are classified. Hard-gate confidence is a qualitative estimate of determinism, not a statistical measure.