ClinicalTrials.gov Document Cover Page

Statistical Analysis Plan

Official Title of the Study:	Prospective Multicenter Observational Evaluation of a Risk-Adapted Surveillance Strategy for First-Site Lung and Bone Metastasis in Hepatocellular Carcinoma
NCT Number:	NCT07444905
Document Date:	January 2, 2022
Document Type:	Statistical Analysis Plan
Protocol ID:	MLEHM-010
Version:	Version 1.0; pre-live frozen operative package
Institution:	Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

The substantive study document begins on the next page.

Prospective Multicenter Observational Evaluation of a Risk-Adapted Surveillance Strategy for First-Site Lung and Bone Metastasis in Hepatocellular Carcinoma

Statistical Analysis Plan

Protocol ID MLEHM-010
Version 1.0; pre-live frozen operative package
Version / effective date January 2022
Institution Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

1. Scope and analysis principles

This Statistical Analysis Plan specifies the principal observational analysis framework for the linked programme of international model transportability, prospective implementation, and comparative observational evaluation of the lung-metastasis and bone-metastasis strategy. No randomisation is performed, and no estimate is to be interpreted as a randomised treatment effect.

- Freeze principle: predictors, preprocessing, model specification, thresholds, SOP-linked pathways, and principal implementation-analysis objects are fixed before first live implementation and carried forward unchanged during the primary implementation phase.
- Wave 2 uses the same v1.0 package without retraining, deployment recalibration, threshold revision, or SOP remapping.
- Displayed cohort summaries are unweighted descriptive values unless explicitly stated otherwise; weighted diagnostics are reported separately.
- The prospective observational programme is planned to start in February 2022 and is not restricted to postoperative or curative-intent-only populations.
- The primary outcome for the prospective implementation analysis is stratum-specific pathway-action fulfilment within 60 days after the eligible index assessment, defined as completion or formal scheduling/documentation by day 60 of the stratum-specific prespecified pathway action assigned at that assessment. Secondary implementation, pathway-linked detected-event, and key patient-important outcomes are specified below. Incident EHM and non-target EHM events are retained for disease-event ascertainment, competing-risk classification, descriptive clinical context, and site-specific model-evaluation context. Overall survival and RMST are supportive long-term outcomes from the prespecified day-60 landmark and are not the primary causal claim.

2. Principal prespecified analyses versus supportive analyses

Analytical domain	Interpretive role	Prespecified status
Performance and transportability	Principal prespecified	12-month discrimination and calibration at prespecified horizons across held-out, international, and prospective settings.
Stratum-specific pathway-action fulfilment within 60 days	Primary implementation outcome	Core prospective implementation analysis within day 60, using completion or formal scheduling/documentation of the assigned stratum-specific pathway action.
Time to first prespecified pathway action	Secondary implementation outcome	Timeliness analysis within the day-60 implementation window.
Downstream management activation	Secondary implementation / mechanistic outcome	Mechanistic implementation outcome in the full eligible cohort.
Asymptomatic actionable detection, symptom-driven detection, and acute deterioration	Key patient-important secondary outcomes	Core patient-important comparative outcomes through 12 months.
Post-detection treatment-changing action	Pathway-linked detected-event outcome	Supportive mechanistic analysis among detected qualifying LM/BM events.
Landmark overall survival and RMST	Supportive long-term outcomes	Long-term outcome summaries from the day-60 landmark; not the primary causal claim.
ITS, staggered difference-in-differences, event-study, placebo timing, negative-control endpoint	Supportive corroboration	Quasi-experimental or falsification analyses that do not supersede the principal observational comparator framework.

3. Analysis populations

Population	Definition	Size status
Development cohort	Domestic retrospective cohort for model development and held-out internal evaluation.	Fixed pre-implementation cohort; n=15,688.
International external-evaluation cohort	Independent held-out retrospective cohort for external transportability.	Fixed pre-implementation cohort; n=4,451.
Wave 1 implementation cohort	Prospective implementation cohort at 11 domestic centres.	All eligible accrued episodes during wave 1 according to this SAP.
Wave 2 implementation cohort	Independent prospective new-centre replication at 4 domestic centres from a distinct health-system setting within the same national context.	All eligible accrued episodes during wave 2 according to this SAP.
Pooled prospective cohort	All prospective implementation episodes across both waves and across routine treatment pathways represented at time zero.	Final analytic size determined after data lock.
Eligible strategy-period comparative cohort	Prospective eligible index episodes under the prespecified strategy across routine treatment pathways.	Final analytic size determined after data lock.
Aligned usual-care comparator	Centre- and calendar-epoch-aligned eligible usual-care episodes.	All episodes meeting comparator mapping rules.

4. Time anchors, exposure definitions, and follow-up

Anchor / definition	Operational specification
Time zero	The eligible index assessment, defined as the first post-diagnostic, pre-treatment surveillance-planning assessment for a strategy-period episode or the corresponding centre- and calendar-epoch-aligned usual-care decision point.
Implementation grace period	Day 0 through day 7 after time zero.
Primary implementation window	Day 0 through day 60.
Patient-important detection window	Day 0 through 12 months after time zero.
Long-term survival window	From the day-60 landmark through 36 months.
Primary observational exposure	Eligible strategy-period episodes versus centre- and calendar-epoch-aligned usual-care episodes; contemporaneous eligible no-tool episodes retained as sensitivity only.

5. Endpoint structure and estimands

Domain	Endpoint	Primary scale
Performance	12-month lung- and bone-metastasis discrimination and calibration.	tdAUC; calibration intercept and slope; decision-range calibration.
Implementation - primary	Stratum-specific pathway-action fulfilment within 60 days.	Adjusted risk ratio and adjusted absolute difference.
Implementation	Time to first prespecified action.	Adjusted hazard ratio and adjusted absolute difference in days.
Implementation	Downstream management activation within 60 days (full eligible cohort).	Adjusted risk ratio and adjusted absolute difference.
Implementation	Post-detection treatment-changing action within 60 days (qualifying detections).	Adjusted risk ratio and adjusted absolute difference.
Implementation	Strategy-nonconcordant intensified action within 60 days.	Adjusted risk ratio and adjusted absolute difference.
Patient-important	Asymptomatic actionable detection; symptom-driven detection; ED / hospitalisation; acute deterioration; actionability at discovery; BM skeletal-related events.	Adjusted risk ratio or hazard ratio with adjusted absolute differences.
Disease-event / model context	First LM and first BM; incident EHM and non-target EHM for event ascertainment and competing-risk classification.	Cumulative incidence and descriptive / model-evaluation summaries.
Supportive long-term	Landmark overall survival, RMST gain, supportive fixed-time milestones, and subgroup heterogeneity.	Adjusted hazard ratio; absolute survival difference; RMST gain; interaction tests.

For the primary implementation endpoint, the same locked stratum-specific risk-to-action mapping is applied in strategy-period and aligned usual-care episodes. Qualifying fulfilment comprises documented continuation or scheduling of routine review for low-risk episodes, earlier focused reassessment or formal scheduling/documentation for intermediate-risk episodes, and expedited LM/BM-directed imaging, MDT review, referral, or formal scheduling/documentation thereof for high-risk episodes. A vague undocumented intention to follow up as needed is not sufficient; the qualifying record must be dated on or before day 60 and traceable to clinic documentation, scheduling systems, imaging orders/bookings, MDT logs, referral records, treatment-planning records, or pathway documentation.

6. Prespecified pre-exposure covariates

The weighting and balance framework uses prespecified pre-exposure covariates measured at time zero. These include sex, age, liver disease aetiology, antiviral therapy, BCLC stage, largest tumour diameter, multiplicity, portal vein tumour thrombus, ECOG performance status, Child-Pugh class, ALBI grade, AFP, and other baseline variables available in the implementation dataset. Covariate balance is summarised using absolute standardised mean differences, with an absolute SMD below 0.10 considered acceptable.

7. Performance and transportability analyses

- Discrimination is summarised using time-dependent area under the receiver-operating characteristic curve at prespecified horizons, including 12 months.
- Calibration is summarised using intercept and slope, together with observed 12-month incidence within prespecified operational bands around the thresholds.
- Threshold stability and monitoring review are assessed across the international external-evaluation cohort, wave 1, and wave 2 without updating the package during the primary implementation phase.
- Prespecified subgroup transportability is described across region, aetiology, planned treatment or management context, and BCLC stage.

8. Comparative implementation analyses

- Marginal comparative estimates are estimated using inverse-probability weighting with centre-clustered robust standard errors.
- Stratum-specific pathway-action fulfilment, downstream management activation, and post-detection treatment-changing action are analysed using adjusted risk ratios and adjusted absolute differences in the relevant denominators.
- Time to first prespecified action is analysed using time-to-event methods within the day-60 window and summarised as an adjusted hazard ratio and adjusted absolute difference in days.
- Strategy-nonconcordant intensified action is evaluated among low- and intermediate-risk episodes only; symptom-driven overrides are excluded from this endpoint by design.

9. Patient-important comparative analyses

- Asymptomatic actionable lung- or bone-metastasis detection and symptom-driven detection are evaluated within 12 months after time zero.
- Metastasis-related emergency-department visit or hospitalisation and the acute-deterioration composite are evaluated within 12 months after time zero.

- Detection while still clinically actionable at discovery is restricted to detected events with adjudicable actionability status at first discovery.
- The BM skeletal-related event composite includes pathologic fracture, spinal cord compression, urgent decompression/stabilisation, or urgent palliative radiotherapy for BM within 12 months.

10. Supportive long-term survival analyses

- Overall survival is analysed from the day-60 landmark through 36 months.
- The principal supportive long-term package includes adjusted landmark survival curves, 36-month survival percentages, adjusted hazard ratio, absolute survival difference, and restricted mean survival time through 36 months.
- A supportive fixed-time 24-month survival milestone is retained as a secondary summary.
- Predefined subgroup analyses include the high-risk subgroup and interaction between strategy exposure and the prespecified baseline risk strata.
- These observational landmark estimates are supportive and are not interpreted as causal survival-effect estimates.

11. Quasi-experimental support and falsification

- Segmented interrupted time-series, staggered difference-in-differences, and event-study analyses are used to evaluate temporal corroboration of implementation-period estimates.
- Placebo timing analyses, pre-period placebo checks, and a prespecified negative-control endpoint are included as falsification components when data availability permits.
- The prespecified negative-control endpoint is non-metastasis-related emergency-department visit or unplanned hospitalisation within 60 days.
- These analyses are supportive and do not supersede the principal observational comparative framework.

12. Missing data, censoring, and competing events

Missingness, administrative censoring, loss to follow-up, and competing events are handled according to the operational analysis rules. Row-specific denominators may differ across endpoints because of endpoint-specific analytic subsets, adjudication availability, patient-level collapsing, and 60-day landmarking. Continuous covariates summarised as medians and interquartile ranges are modelled on their original analysis scale where required.

13. Software, reproducibility, and reporting

Analyses are conducted using version-controlled scripts under R and Python. Reproducibility is supported through version-controlled code, standard quality-control procedures, and institutional data-governance requirements.

14. Companion documents

This SAP should be read together with the study protocol, endpoint charter, and governance note.