Statistical Analysis Plan

Study	**Clinical characteristics and CRI-defined high-risk presentation among adults evaluated for anti-glomerular basement membrane antibodies: a retrospective cross-sectional analytical study at Baoding No.1 Central Hospital**
SAP version/date	Version 1.0 - 28 September 2026
Analysis software	IBM SPSS Statistics 29.0
Final analytical sample	91 unique index records: 64 anti-GBM-positive and 27 anti-GBM-negative
Primary analysis cohort	Anti-GBM-positive cohort for CRI risk-class analyses
Primary outcome	CRI-defined high-risk versus lower-risk presentation at index evaluation
Study design	Retrospective cross-sectional analytical study

1. Analysis principles

Analyses are cross-sectional and use the first eligible index record for each patient. Statistical tests are two-tailed with alpha 0.05. Effect estimates and 95% confidence intervals are emphasized. The CRI outcome is a rule-based classification, so variables that define the outcome are not modelled as independent predictors of that same outcome.

2. Analysis populations

1. Source cohort: all 91 eligible index records, including anti-GBM-positive and anti-GBM-negative patients.
2. CRI cohort: the 64 anti-GBM-positive patients, classified as high risk or lower risk.
3. Biopsy-specific analyses: patients with a verifiable renal biopsy result for the variable being examined.

3. Variable definitions

Anti-GBM positivity is >=20 RU/mL. Anti-GBM titer is categorized as low (20 to <100 RU/mL) or high (>=100 RU/mL). CRI high-risk presentation is defined by at least one of: eGFR <15 mL/min/1.73 m², dialysis dependence, pulmonary hemorrhage, or anti-GBM/ANCA co-positivity. The primary outcome is measured once at index presentation. Pulmonary hemorrhage is determined using compatible clinical and imaging evidence. eGFR is calculated using the 2021 CKD-EPI creatinine equation.

4. Descriptive analysis

Normally distributed continuous variables are summarized by mean ± standard deviation; skewed variables by median [interquartile range]; categorical variables by n (%). Normality is assessed using the Shapiro-Wilk test and variance homogeneity using Levene's test where applicable.

5. Source-cohort comparisons

Baseline demographic, clinical, laboratory, serologic, imaging, and biopsy features are compared between anti-GBM-positive and anti-GBM-negative patients using independent-samples t tests for approximately normal continuous variables, Mann-Whitney U tests for skewed continuous variables, and Pearson chi-square or Fisher exact tests for categorical variables.

6. Primary outcome analysis

The number and percentage of anti-GBM-positive patients classified as CRI high risk and lower risk are reported with the prevalence of each constitutive criterion. Because eGFR <15 mL/min/1.73 m^2, dialysis dependence, pulmonary hemorrhage, and anti-GBM/ANCA co-positivity define the outcome, statistical associations involving those variables are interpreted only as descriptive properties of the classification and not as independent predictor evidence.

7. Secondary outcome analyses

CRI high-risk and lower-risk groups are compared on renal, pulmonary, serologic, and histopathologic secondary variables using t tests, Mann-Whitney U tests, chi-square tests, or Fisher exact tests as appropriate. Effect estimates, group summaries, and exact p values are reported. Missing denominators for biopsy-specific variables are stated where applicable.

8. Exploratory association analyses

Exploratory univariable binary logistic regression evaluates non-defining baseline features in relation to CRI high-risk classification. Prespecified candidate variables are age, blood urea nitrogen, serum albumin, high anti-GBM titer (>=100 RU/mL), crescent burden >=50%, glomerulosclerosis, and tubular injury. Results are reported as odds ratios with 95% confidence intervals and p values. No multivariable predictor model is used because the outcome is constructed from clinical severity variables and the positive cohort is modest in size. These analyses are hypothesis-generating and do not establish causation or prognostic performance.

9. Prespecified subgroup summaries

1. Symptomatic versus asymptomatic anti-GBM-positive presentation.
2. Single-positive anti-GBM versus double-positive anti-GBM/ANCA serology.
3. Low anti-GBM titer (20 to <100 RU/mL) versus high titer (>=100 RU/mL).
4. Pulmonary hemorrhage present versus absent, reported descriptively because pulmonary hemorrhage is a CRI-defining criterion.

10. Correlation analyses

Spearman rank correlation may be used to describe associations between anti-GBM titer and continuous markers such as serum creatinine, eGFR, or crescent burden where data are available. Correlations are exploratory.

11. Missing data

Eligibility requires availability of the core variables needed for CRI classification. For secondary variables, analyses use available-case denominators and the denominator is stated when it differs from the full group. No outcome values are imputed. Reasons for record exclusion are summarized in the study flow.

12. Data quality and sensitivity

Only verifiable source records are included. Duplicate encounters are removed and only the first eligible index presentation is retained. Uncertain categorical abstraction and CRI assignments undergo dual review. Cohen kappa is used to assess agreement for categorical abstraction, with a prespecified acceptance threshold of >=0.80.

13. Multiplicity and interpretation

No formal multiplicity adjustment is applied. Secondary comparisons and exploratory regression analyses are interpreted in the context of the number of analyses, the modest sample size, and the cross-sectional design. CRI is not evaluated as a longitudinal prognostic score in this study.

14. Reporting

Participant/record flow, baseline characteristics, CRI risk distribution, secondary feature comparisons, exploratory associations, and relevant subgroup summaries are reported. Results are presented without implying temporal causation. Any future predictive use of CRI requires external prospective validation.