STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1	To investigate the effects of different storage conditions on the stability of human cytokeratin 18-M30 (CK18-M30) in serum
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1	A total of 22 serum samples from individuals undergoing routine physical examination were collected. Baseline value from the immediate test. The remaining samples were distributed into five groups. CK18-M30 is relatively stable at -20°C. Between 2 and 8°C, the separation gel tubes were more stable than the aliquot tubes for up to three days. Aliquot tubes stored at 2-8℃ were suboptimal, as the CK18-M30 became unstable after 3 days.
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	1	The level of CK18-M30 has been used as a diagnostic and prognostic biomarker for acute and chronic liver diseases, and it correlates with histological inflammatory activity (Erkisa et al., 2021). Serum or plasma levels are widely used better to define cell death in various forms of liver disease (Cebi et al., 2024). Serum cytokeratin 18-M30 has become a significant biomarker in the pharmaceutical sciences for

				assessing hepatocellular apoptosis associated with medication exposure (Fujihara., 2025). It is being used more and more in clinical trials, pharmacovigilance initiatives and drug development programs to help with the early identification and tracking of drug-induced liver damage. Due to logistical limitations, serum samples are frequently gathered from various locations and examined after varying storage times in these pharmacological contexts. Therefore, the primary aim of this study was to systematically evaluate the short-term stability of serum CK18-M30 under different storage temperatures, durations and tube types commonly encountered in clinical laboratory practice. Proper pre-analytical sample handling, such as specimen collection, processing and storage are very important. To date, however, none of them has been studied with respect to their stability.
Objectives	3	State specific objectives, including any prespecified hypotheses	2	The goal of this study is to determine how consistent the serum CK18-M30 test results are across different sample tubes, storage duration, and storage temperature.
Methods				
Study design	4	Present key elements of study design early in the paper	2	A total of 22 serum samples with varying CK18-M30 concentrations were randomly

				collected from individuals undergoing routine physical examination with no documented acute or chronic liver disease at The Fourth Affiliated Hospital, Zhejiang University School of Medicine in 2025. The hospital's institutional review board approved all procedures.
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	2	A total of 22 serum samples with varying CK18-M30 concentrations were randomly collected from individuals undergoing routine physical examination with no documented acute or chronic liver disease at The Fourth Affiliated Hospital, Zhejiang University School of Medicine in 2025. The hospital's institutional review board approved all procedures. All selected venous blood specimens were centrifuged at 2,000g for 10 min. An aliquot of serum was immediately assayed (within 1 hr post - centrifugation) to determine baseline CK18-M30 concentrations. All baseline measurements were performed within a single analytical run to minimize inter-run variability. Stored samples were analyzed in subsequent runs, using the same analyzer, identical calibration settings, and reagents from a single lot. Quality control materials were

				included in each run to ensure analytical consistency, and baseline samples were not re-measured alongside stored samples. The remaining serum was evenly divided into five storage groups (300 ul each): Group A: Stored at -20°C for 3 days in polypropylene aliquot tubes. Group B: Stored at 2–8°C for 3 days in polypropylene aliquot tubes. Group C: Stored at 2–8°C for 3 days in inert separation gel tubes (BD Vacutainer®, BD Company). Group D: Stored at 2–8°C for 7 days in inert separation gel tubes (BD Vacutainer®, BD Company). Group E: Stored at 2–8°C for 7 days in polypropylene aliquot tubes. At 3 or 7 days, the samples were collected and incubated at RT for 30 min before immediate assay. Results were recorded and compared against baseline for stability per condition.
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	2	A total of 22 serum samples with varying CK18-M30 concentrations were randomly collected from individuals undergoing routine physical examination with no documented acute or chronic liver disease at The Fourth Affiliated Hospital, Zhejiang University School of Medicine in 2025. The hospital's institutional review board approved all procedures.
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and		Not applicable

		unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case		
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	2	All selected venous blood specimens were centrifuged at 2,000g for 10 min. An aliquot of serum was immediately assayed (within 1 hr post - centrifugation) to determine baseline CK18-M30 concentrations. All baseline measurements were performed within a single analytical run to minimize inter-run variability. Stored samples were analyzed in subsequent runs, using the same analyzer, identical calibration settings, and reagents from a single lot. The remaining serum was evenly divided into five storage groups (300 ul each): Group A: Stored at -20°C for 3 days in polypropylene aliquot tubes. Group B: Stored at 2–8°C for 3 days in polypropylene aliquot tubes. Group C: Stored at 2–8°C for 3 days in inert separation gel tubes (BD Vacutainer®, BD Company). Group D: Stored at 2–8°C for 7 days in inert separation gel tubes (BD Vacutainer®, BD Company). Group E: Stored at 2–8°C for 7 days in polypropylene aliquot tubes. At 3 or 7 days, the samples were collected and incubated at RT for 30 min before immediate assay. Results were recorded and compared against baseline

				for stability per condition.
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	2	The primary instruments used in this trial included an automated chemiluminescence immunoassay analyzer (C3000; Hotgen), a laboratory centrifuge (ThermoFisher Scientific), a 2–8°C refrigerated storage unit (MPC-5V1006; Zhongkeduling), and a -20°C ultra-low freezer (MDF-25V2268E; Zhongkeduling). The assay reagents comprised biological CK18-M30 detection reagents, calibrators, and quality control materials (all from Hotgen). Furthermore, sample collection and processing were performed using VACUETTE 3.5 mL serum tubes containing inert separation gel and 1.5 mL disposable polypropylene aliquot tubes (Kangjian). CK18-M30 concentrations were measured using the Hotgen C3000 automated chemiluminescence immunoassay analyzer (Hotgen Biotech Co., Ltd., Beijing, China). The assay is based on a sandwich chemiluminescent immunoassay principle using monoclonal antibodies specific for caspase-cleaved CK18 (M30 epitope). The assay was calibrated according to the manufacturer's instructions prior to analysis. Internal quality control materials at two concentration levels were analyzed in each run. The

				analytical measuring range, limit of detection, and calibration frequency followed the manufacturer's specifications.
Bias	9	Describe any efforts to address potential sources of bias	2	All baseline measurements were performed within a single analytical run to minimize inter-run variability. Stored samples were analyzed in subsequent runs, using the same analyzer, identical calibration settings, and reagents from a single lot. Quality control materials were included in each run to ensure analytical consistency, and baseline samples were not re-measured alongside stored samples. All assays were run on the Hotgen C3000 analyzers, and each run included QC samples. In the laboratory's internal quality control system, the CV is 3%, which is much lower than the manufacturer's inter-batch CV threshold of less than 15%. Experimental samples were submitted for analysis only when the IQC results met the required criteria, i.e., a CV $\leq$ 3%. This strict process ensured the results could be trusted and reproduced—laboratory environment followed by the manufacturer’s guideline of the Hotgen C3000 analyzer. The analyzer ran without any error alarms during the test, as confirmed by real-time system log monitoring. All assays used

				reagents from a single lot (CK18-M30 detection reagent, Lot No. C24060111) to reduce lot-to-lot variation. A standardized approach was used to tightly control pre-analytical and analytical variables, thereby increasing the reliability of comparative stability between storage groups.
Study size	10	Explain how the study size was arrived at	2,4,5	A total of 22 serum samples with varying CK18-M30 concentrations were randomly collected from individuals undergoing routine physical examination with no documented acute or chronic liver disease at The Fourth Affiliated Hospital, Zhejiang University School of Medicine in 2025. First, the study included only 22 serum samples collected from individuals undergoing routine physical examination at a single center. Although this number is acceptable for a preliminary stability assessment, the limited sample size may restrict the generalizability of the findings to broader clinical trial or pharmacovigilance settings.

Continued on next page

Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	2	All selected venous blood specimens were centrifuged at 2,000g for 10 min. An aliquot of serum was immediately assayed (within 1 hr post - centrifugation) to determine baseline CK18-M30 concentrations. All baseline measurements were performed within a single analytical run to minimize inter-run variability. Stored samples were analyzed in subsequent runs, using the same analyzer, identical calibration settings, and reagents from a single lot. At 3 or 7 days, the samples were collected and incubated at RT for 30 min before immediate assay. Results were recorded and compared against baseline for stability per condition.
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	2,3	The data were analyzed with SPSS Statistics 29. The Shapiro-Wilk test showed that cytokeratin 18-M30 concentrations were not normally distributed in all groups, so the results were presented as median (IQR). Friedman's test was performed across all groups to determine whether there were overall differences, and pairwise Wilcoxon Signed-Rank tests were conducted for each group compared to the control. Statistical significance was set at $P < 0.05$. Because multiple pairwise comparisons were performed, a Bonferroni-adjusted significance threshold was applied to reduce the risk of type I error. The adjusted P-value threshold was calculated based on the number of pairwise comparisons performed.
		(*b*) Describe any methods used to examine subgroups and interactions	3	Groups were considered stable only if they fulfilled both predefined criteria:

				(1) ≥80% concordance with baseline values and (2) no statistically significant difference compared with baseline ($P \geq 0.05$). Because multiple pairwise comparisons were performed, a Bonferroni-adjusted significance threshold was applied to reduce the risk of type I error. The adjusted P-value threshold was calculated based on the number of pairwise comparisons performed. The study's stability acceptance criteria and concordance thresholds align with the bioanalytical validation principles commonly used in clinical medication trials and pharmaceutical research.
		(*c*) Explain how missing data were addressed		Not applicable
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	3	Groups were considered stable only if they fulfilled both predefined criteria: (1) ≥80% concordance with baseline values and (2) no statistically significant difference compared with baseline ($P \geq 0.05$).
		(*e*) Describe any sensitivity analyses	2	Friedman's test was performed across all groups to determine whether there were overall differences, and pairwise Wilcoxon Signed-Rank tests were conducted for each group compared to the control. Statistical significance was set at $P < 0.05$. Relative deviation threshold: Per reagent manufacturer's instructions, post-storage values deemed stable if ≤±15% deviation from baseline. Concordance rate: CNAS - CL02 - A003: 2018 for CL, each storage group should have achieved a concordance rate of at least 80%.
Results				
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and	3	Group A showed no significant difference from baseline ($P > 0.05$).

		analysed		Concordance reached 100%, meeting validation criteria. Group C showed no significant difference from baseline (P > 0.05). With a 90.9% concordance rate, it passed stability assessment. Group B showed a significant difference from baseline (P < 0.05), with a 77.3% concordance rate below the ≥80% threshold. Group D had a significant difference from baseline (P < 0.05), with a 72.7% concordance rate below the required standard. Group E had a significant difference from the baseline group (P < 0.01), with a concordance rate of only 40.9%. The results are shown in Table 1.
		(b) Give reasons for non-participation at each stage		Not applicable
		(c) Consider use of a flow diagram		Not applicable
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	2,3	The data were analyzed with SPSS Statistics 29. The Shapiro-Wilk test showed that cytokeratin 18-M30 concentrations were not normally distributed in all groups, so the results were presented as median (IQR). The Friedman test indicated a significant overall difference among all storage groups (P < 0.001). To further identify specific differences, Wilcoxon signed-rank tests were conducted to compare each experimental group with the baseline group. The results of these tests are as follows: Group A showed no significant difference from baseline (P > 0.05). Concordance reached 100%, meeting validation criteria. Group C showed no significant difference from baseline (P > 0.05). With a 90.9% concordance rate, it passed stability assessment. Group B showed a significant

				difference from baseline (P < 0.05), with a 77.3% concordance rate below the ⩾80% threshold. Group D had a significant difference from baseline (P < 0.05), with a 72.7% concordance rate below the required standard. Group E had a significant difference from the baseline group (P < 0.01), with a concordance rate of only 40.9%. The results are shown in Table 1.
		(b) Indicate number of participants with missing data for each variable of interest		Not applicable
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)		Not applicable
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time		Not applicable
		Case-control study—Report numbers in each exposure category, or summary measures of exposure		Not applicable
		Cross-sectional study—Report numbers of outcome events or summary measures	3	Group A showed no significant difference from baseline (P > 0.05). Concordance reached 100%, meeting validation criteria. Group C showed no significant difference from baseline (P > 0.05). With a 90.9% concordance rate, it passed stability assessment. Group B showed a significant difference from baseline (P < 0.05), with a 77.3% concordance rate below the ≥80% threshold. Group D had a significant difference from baseline (P < 0.05), with a 72.7% concordance rate below the required standard. Group E had a significant difference from the baseline group (P < 0.01), with a concordance rate of only 40.9%. The results are shown in Table 1. Comparison of Group A with baseline, 12 sample results below baseline with relative deviations of 2.1%–14.7%. 10 sample results above baseline with relative deviations of 2.2%–14.3%. 6 sample results >10% deviation (all with baseline CK18-M30 <150 U/L).

				Comparison of Group B with baseline, 14 sample results below baseline with relative deviations of 1.2%–24.5%. 8 sample results above baseline with deviations of 0.1%–8.3%. All 7 samples with >10% deviation were below baseline, with baseline cytokeratin 18-M30 values of 26.6–1,124 U/. No samples above baseline exceeded 10% deviation. Comparison of Group C with baseline, 13 sample results below baseline with relative deviations of 0.1%–22.1%. 9 sample results above baseline with deviations of 0.5%–13.7%. 6 Samples with >10% deviation, only Sample number 10 had a baseline value >500 U/L and the remaining five Samples were <150 U/L. The results are shown in Table 2. Comparison of Group D with baseline,16 sample results below baseline with relative deviations of 0.4%–31.7%, including 6 exceeding the ±15% threshold (all with baseline CK18-M30 <150 U/L).6 sample results above baseline with relative deviations of 0.2%–12.9%, all within the ±15% acceptable range. Comparison of Group E with baseline, 18 sample results below baseline with relative deviations of 2.5%–36.9%, including 12 exceeding 15% (baseline CK18-M30: 26.6–1,124 U/L).4 sample results above baseline with relative deviations of 2.8%–32.3%, with one sample exceeding 15%. The results are presented in Table 3.
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their	3	A comparison of the five storage

		precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included		groups with the baseline values reveals that the absolute deviations for Groups A, B, and C were all close to zero, consistent with their concordance rates. Group A demonstrated the least variability from baseline, while Group E exhibited the greatest absolute deviation and variability (Fig. 1). With relative deviation, Group C demonstrated the closest proximity to zero, while Group A exhibited the least variability in relative deviation. Conversely, Group E exhibited the most substantial relative deviation and variability (Fig. 2).
		(*b*) Report category boundaries when continuous variables were categorized		Not applicable
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period		Not applicable

Continued on next page

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses		Not applicable
Discussion				
Key results	18	Summarise key results with reference to study objectives	4,5,6	As well as being included in the guidelines and treatments from different countries (Kulkarni et al., 2026). The current literature has shown that most laboratory errors occur during the preanalytical process (Wei et al., 2023; Winter et al., 2023). Therefore, it is very important to ensure the quality of the preanalytical stage. Based on previous studies, we selected clinical samples covering low, medium and high concentrations, with more than 20 samples in each category, to reach the medical decision level (Gomez-Rioja et al., 2023). Compared to the baseline, according to the experimental data, the samples stored in aliquot tubes at -20°C for 3 days and those stored in separation gel tubes at 2-8°C for 3 days did not show a statistically significant difference. Experimental data show that the original separation gel tubes are stable at 2-8°C for 3 days or less, thus supporting the storage of untested samples on the same day in the laboratory and can be stored for a longer time while maintaining stability.
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	4,5	However, several limitations should be acknowledged. First, the study included only 22 serum samples collected from individuals undergoing routine physical examination at a single center. Although this number is acceptable

				for a preliminary stability assessment, the limited sample size may restrict the generalizability of the findings to broader clinical trial or pharmacovigilance settings. Therefore, the results should be interpreted primarily as laboratory-level evidence rather than definitive guidance for large-scale multicenter applications. In addition, lower storage temperatures such as -80°C were not evaluated and should be considered in future studies, particularly for long-term biobanking applications.
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	6	From a laboratory perspective, the stability of CK18-M30 is relevant for institutions performing delayed sample analysis or centralized testing (Zhang et al., 2024). The present findings provide practical guidance for hospital laboratories and clinical research units regarding acceptable short-term storage conditions. However, given that this study was conducted at a single center using one analyzer platform and a single reagent lot, extrapolation to regulatory policy or large-scale pharmacovigilance frameworks should be made cautiously. Multicenter validation studies using different analytical systems would be necessary before broader implementation.
Generalisability	21	Discuss the generalisability (external validity) of the study results	6,7	From a laboratory perspective, the stability of CK18-M30 is relevant for institutions performing delayed sample analysis or centralized testing (Zhang et al., 2024). The present findings provide practical

				guidance for hospital laboratories and clinical research units regarding acceptable short-term storage conditions. In conclusion, serum CK18-M30 demonstrated acceptable short-term stability when stored at −20°C for 3 days and in separation gel tubes at 2–8°C for up to 3 days. Storage at 2–8°C in polypropylene aliquot tubes beyond 3 days resulted in significant deviations from baseline values. These findings provide laboratory-level evidence to support appropriate pre-analytical handling of CK18-M30 when immediate analysis is not feasible. Further multicenter studies involving larger and clinically diverse populations are warranted to confirm these observations.
Other information				
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	7	This work was supported by the Department of Clinical Laboratory, the Fourth Affiliated Hospital, Zhejiang University School of Medicine, Yiwu, China.

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.