

# SAP

## Statistical Analysis Plan (SAP) — Study II

### Tele-mentored eFAST in postoperative liver surgery patients

Version	v1.0
Date	2026-05-03
ClinicalTrials.gov	<a href="https://clinicaltrials.gov/ct2/show/study/NCT06594562">NCT06594562</a>
Protocol ID	172964
Privacy approvals (PVO, OUS)	21/03184 (main frame); 25/21566 (AI image reuse with SINTEF, valid 09.08.2025–01.09.2027)
Author / PI	Peder Christian Engelsen, OUS

Changes after the issue date are logged at the bottom (change log) with date and rationale.

### Study design and endpoints

Prospective single-arm diagnostic agreement study. N = 40 postoperative liver surgery patients. Two independent raters per patient, same time point:

- **Index:** nurse performs tele-mentored eFAST (mentored by radiologist via video link).
- **Reference:** experienced radiologist performs independent bedside eFAST within one hour.

Agreement interpretations are conditional on the radiologist's bedside findings as the best available standard.

### Registered primary outcomes (ClinicalTrials.gov NCT06594562)

1. **Concordance rate (p\_o)** — percent agreement between telementored examiner and radiologist, per window and pooled abdomen.
2. **Exam duration** — time from probe contact to agreement between expert and examiner. Reported as median (IQR).
3. **Image quality** — per window and overall. Radiologist rating on Likert 1–5 per window.

## **Registered secondary outcomes**

- Cognitive load: NASA Task Load Index (1–9) per exam.
- User experience: questionnaire.
- Learning curve: NASA-TLX and exam duration over sequence number.

# Agreement methodology

## Window structure — pre-specified:

Analysis	Construction
Abdomen (primary)	Pooled composite: OR of RUQ, LUQ, suprapubic on each side (tele and gold). A patient is "abdomen-positive" if $\geq 1$ window shows free fluid.
Pericardium (primary)	1:1, tele vs gold. Fluid is anatomically confined.
Pneumothorax / lung (primary)	1:1, tele vs gold.
RUQ, LUQ, suprapubic (secondary / transparency)	Per-window $\kappa$ to show what pooling obscures.

Pooling is clinically justified (free peritoneal fluid migrates with patient position).

## Metrics per primary analysis:

Metric	Role
p_o (observed agreement)	Registered primary endpoint.
Cohen's $\kappa$ (unweighted)	Secondary. Agreement beyond chance. Landis & Koch (1977): $<0$ worse than chance, 0.00–0.20 poor, 0.21–0.40 slight, 0.41–0.60 moderate, 0.61–0.80 substantial, 0.81–1.00 almost perfect.
Gwet's AC1	Secondary. Prevalence-robust alternative to $\kappa$ . Reported for all primary analyses; most informative at low prevalence (pericardium, lung).
Sensitivity, specificity, PPV, NPV (Wilson 95 % CI)	Secondary. Reported with the radiologist as "reference standard" and explicitly framed as an agreement surrogate, not absolute diagnostic accuracy.

## Confidence intervals for agreement metrics:

Not reported in the first iteration of SAP v1.0. Point estimates are given with explicit discussion of sampling variability at  $n = 40$ . Bootstrap BCa 95 % CI (2000 resamples, patient-level) may be added in the manuscript version without a SAP revision.

## Success thresholds (pre-specified interpretation)

Endpoint	Clinically acceptable threshold	Rationale
p_o abdomen pooled	$\geq 80 \%$	Conventional "substantial agreement"; comparable to historical FAST data (Rozycki 1998 reported ~86 % sensitivity for experienced operators).
p_o pericardium	$\geq 90 \%$	Higher threshold due to low prevalence.
p_o lung / PTX	$\geq 90 \%$	Same reasoning as pericardium; very low prevalence expected in postop liver surgery.
Exam duration (median)	$\leq 10 \text{ min}$	Acceptable in the postop ward.
Image quality	Median $\geq 3$ per window (adequate for diagnostic interpretation)	McBeth (2022), Likert point 3 = "fair".

Thresholds are used for descriptive interpretation, not hypothesis tests. If an endpoint falls short of its threshold, it is reported as not achieved with discussion.

## Missing data handling

- **Aborted exam:** excluded from analysis.
- Exclusions are reported (lack of consent, termination).

## Stopping rule

- Teleultrasound analysis stops at  $n = 40$ .

## Secondary / exploratory analyses (pre-specified)

1. **Per-window Cohen's  $\kappa$  for RUQ, LUQ, suprapubic.**
2. **Learning curve:** linear regression of NASA-TLX and exam duration on sequence number per nurse.

## Reporting framework

STARD 2015 (Standards for Reporting of Diagnostic Accuracy Studies) is used as a template for manuscript structure.

# Change log

Version	Date	Change
v1.0	2026-05-03	First SAP. Written during ongoing data collection ( $n \approx 26/40$ at time of writing) — not a pre-data-collection SAP. SAP versioning is independent of protocol versioning (protocol currently at v3.1).