

Ensuring Reliability of Medical Devices through Statistical Oversight in the Age of AI

Parveen Kumar | Director- Statistics | Geninvo



Introduction

Core thesis

AI/ML can flag early device drift/failure, but it must be embedded in prespecified statistical oversight to be clinically credible and regulator-ready.

Methods backbone

- Survival analysis (SA): durability / time-to-failure endpoints
- Statistical process control (SPC): in-study drift/shift detection on telemetry
- Automated data-quality (DQ) monitoring: missingness, latency, plausibility, signal quality

Operational scaffold

A 6-pillar lifecycle framework links protocol/SAP/QMS expectations: governance, process, fit-for-purpose validation, traceability, ongoing monitoring, and controlled change management.

Why reliability is harder in connected devices

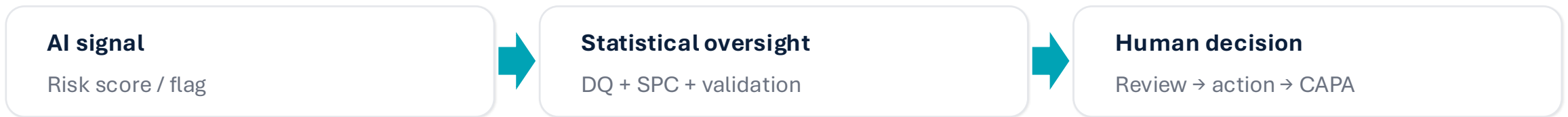
Modern devices generate continuous, high-frequency telemetry in trials and routine care.

- Device malfunction or drift can create patient-safety risk and disrupt drug/device evidence generation.
- Telemetry pipelines introduce non-random missingness (connectivity, non-use, sensor degradation).
- Data-quality issues can bias endpoints, inflate false alerts, and degrade AI model performance.
- Reliability oversight must be designed upfront (protocol + SAP + risk-based monitoring plan).

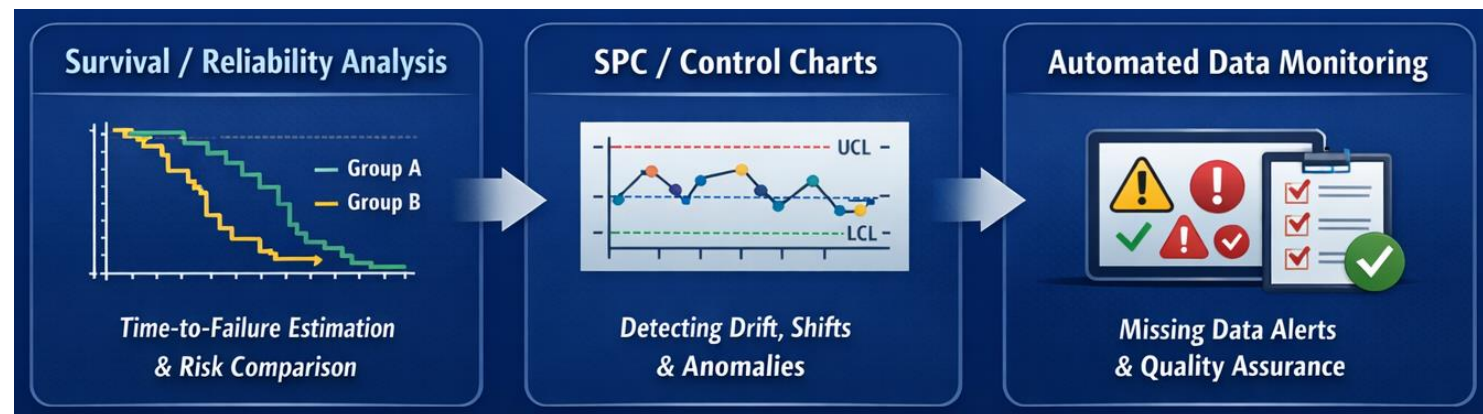


AI/ML as early warning — but not standalone

- AI/ML is good at detecting subtle multivariate patterns preceding malfunction (anomaly detection, drift).
- To be regulator-ready, alerts need: prespecified thresholds, quantified error tradeoffs, and audit trails.
- AI outputs should augment—not replace—clinical/engineering/statistical judgement.



Key idea: prespecified governance turns alerts into defensible evidence.



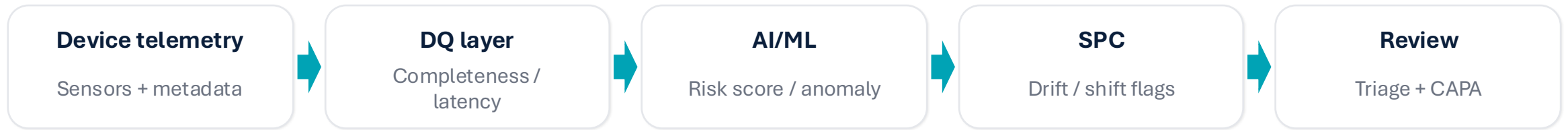
Evidence generation (GCP + estimands)

- Fit-for-purpose, validated systems; data integrity and auditability (ICH E6(R3))
- Pre-specification and robust inference; handling intercurrent events and missing data (ICH E9(R1))
- Risk-based quality management and documentation of decisions

AI-enabled devices / DHTs

- Remote DHT data acquisition: validation + operational controls to reduce missingness (FDA DHT guidance)
- Good machine learning practice (IMDRF GMLP principles)
- Controlled model evolution using Predetermined Change Control Plans (FDA PCCP guidance)
- Alignment across lifecycle approaches (e.g., EMA reflection paper; FDA/EMA 2026 principles)

Hybrid evidence model: AI + statistical oversight- End to end pipeline



What stays prespecified (SAP-friendly)

- Metrics and operating characteristics (error rates, calibration, CIs)
- Control-limit logic and alerting rules (SPC/DQ)
- Endpoint definitions for time-to-failure and intercurrent events

What stays governed (QMS-friendly)

- Human review ownership and escalation pathways
- CAPA documentation and audit trails
- Controlled updates / comparability checks (PCCP)

Output: earlier detection of device issues without compromising endpoint inference.

Six-pillar lifecycle framework

Overview

A prespecified, risk-based scaffold linking AI outputs to protocol/SAP/QMS artifacts.

1. Context-of-use & governance

Intended use, decision rights, escalation paths

2. Operational process

Data flow, roles (RACI), SOPs, CAPA workflow

3. Fit-for-purpose validation

Reference standard + prespecified operating characteristics

4. Traceability & provenance

Auditability from raw signal → decision; version stamping

5. Ongoing monitoring

SPC + data integrity + drift surveillance

6. Change control (PCCP)

Controlled updates, comparability, revalidation triggers

Pillars 1–2: Governance and operational process

Make alerts actionable and auditable

1) Context-of-use & decision governance

- Define the purpose of AI outputs (screening vs decision support)
- Prespecify the error tradeoff (false alerts vs missed true issues)
- Assign review ownership, escalation paths, and decision rights
- Ensure decisions are documented (who/when/why/action)

2) End-to-end operational process

- Define data flow and expected frequency/windows
- Specify RACI, SOPs, and review cadence
- Link alerts to a closure workflow (triage → query → disposition)
- Embed CAPA-style remediation where applicable

Pillar 3: Fit-for-purpose validation

Prespecify operating characteristics

Separate two validations:

- Device measurement validity (sensor accuracy / robustness)
- AI output validity (risk score / classification / anomaly flag)

Minimum validation package (recommended for SAP appendix)

- Reference standard + adjudication rules for true events / false alerts
- Partitioning to prevent leakage (subject-level split; stratify by site/device version)
- Targets + uncertainty: sensitivity, false-alert rate, PPV/NPV, calibration (with CIs)
- Drift metrics + triggers for investigation/revalidation; version-stamping

Pillar 4: Documentation, traceability & provenance

Audit-ready chain of evidence

- Version-stamp device hardware/firmware + model/thresholds + preprocessing/feature specs.
- Maintain source-to-output lineage: raw telemetry → derivations → AI output → decision.
- Log every alert, review outcome, action, and disposition (who/when/why).
- Ensure reproducibility (locked specs/code; rerun capability; archived outputs).

Traceability variables (examples)

- Device/site/version IDs
- Timestamp of trigger
- Rule violated (DQ/SPC)
- Query/issue ID
- Reviewer + outcome
- Action taken
- Closure date/status

Pillar 5: Ongoing statistical monitoring

SPC + data integrity + drift

What to monitor (typical strata)

Participant, device, site, lot/batch, device version/firmware, time window

Metrics: missingness, latency, plausibility flags, signal quality indices, error-code frequency, score distributions

Prespecify alert logic

- Baseline period (“in control”)
- Control limits / run rules
- Re-baselining policy
- Target alert rate (avoid operational noise)

Governance pathway

- Dashboard cadence + owners
- Escalation thresholds
- Investigation steps
- CAPA documentation + closure

Survival analysis (SA) for time-to-failure

Endpoint-level reliability inference

- Define failure in trial-relevant terms: first malfunction, revision/explant, loss of performance beyond threshold, device-related SAE.
- Handle censoring and unequal follow-up naturally (withdrawal, end of study, non-device removals).
- Estimate decision-relevant quantities: event-free probability at timepoints, hazard ratios, median time-to-event (if estimable).
- Use Cox or parametric reliability models (e.g., Weibull) when appropriate; predefine intercurrent event handling.

Typical SA outputs

- Kaplan–Meier curves
- Cumulative incidence
- Cox model estimates
- Weibull shape insights
- Remaining useful life views

Statistical process control (SPC) for telemetry

Early warning on drift/shifts

When to use which chart (rule-of-thumb)

Shewhart: abrupt excursions | EWMA/CUSUM: gradual drift and small persistent shifts

Telemetry signals (examples)

- Injection force / delivered volume
- Temperature drift / calibration index
- Connectivity dropouts / latency
- Error-code frequency
- AI score distribution shifts

Prespecify in protocol/SAP

- Monitoring level (subject/site/version/lot)
- Baseline window + in-control assumptions
- Alerting rules (limits, runs, EWMA/CUSUM thresholds)
- Escalation + CAPA closure criteria

SA vs SPC: keep roles explicit

Dimension	Survival analysis (SA)	SPC
Primary purpose	Durability / time-to-event reliability endpoints	Detect drift/shifts in telemetry for early warning
Data structure	One (or few) event times per subject; censoring central	Many repeated measurements; sequential behavior
Decision cadence	Scheduled analyses per SAP (interim/final)	Frequent/near real-time review with governed response

Guardrail: don't let alerting logic morph into confirmatory endpoint analysis.

Automated data-quality monitoring

Prevent missingness and reduce noise

Why it matters

In connected-device settings, missingness is often informative (connectivity, non-use, sensor malfunction).

Automated DQ monitoring protects downstream AI and endpoint analyses.

Completeness

Missing expected transmissions

Gaps in daily uploads

Incomplete sessions

Timeliness (latency)

Uploads beyond acceptable window

Sync failures / backlogs

Pipeline outages

Plausibility / consistency

Out-of-range values

Impossible sequences

Status contradictions

Signal quality

Artifact-laden waveforms

Uninterpretable records

Low confidence indices

Prespecify thresholds + responses; log trigger → query → resolution for auditability.

Pillar 6: Controlled change management (PCCP-ready)

Manage updates without breaking interpretability

- Prespecify what may change (device firmware, model weights, thresholds, feature definitions).
- Define approvers/governance and evidence required before deployment (impact assessment).
- Maintain comparability: version-stamped datasets; stratified reporting; sensitivity analyses.
- Define revalidation triggers and rollback criteria.

PCCP alignment (high level)

- Predetermined changes (what)
- Methodology to implement (how)
- Impact assessment (evidence)
- User transparency + documentation

Thank you

Contact

Parveen Kumar

Director-Statistics

Geninvo

Email: parveen.k12@geninvo.com

Phone: +91 9871819209