SIIM CAIMI26 Conference on Artificial Intelligence in Medical Imaging
OCT 26-27 | University of Pennsylvania, Philadelphia, PA

The Invisible Feedback Loop: Quantifying Uncompensated Site Labor in Post-Market Evaluation of Diagnostic AI

Inga Edwin, MD, Valley Radiology and Methodist University Cape Fear Valley Medical School

Introduction/Background
Real-world performance data for diagnostic AI most often originates at early deployment sites. The labor required to generate this data is excluded from purchase agreements, untracked operationally, and absent from accounts of post-market AI monitoring. We sought to quantify this labor before proposing formal recognition.

Methods/Intervention
Over five months, a private radiology practice evaluated an FDA-cleared prostate MRI CAD tool deployed on premises. Using dated correspondence, meeting records, and case logs, we reconstructed all work products generated to characterize real-world performance and estimated radiologist time using category-based time bands. Each work product was classified across five attributes: contractual scope, compensation, research-protocol coverage, return of performance data to the site, and attribution in vendor-facing outputs (Figure 1).

Results/Outcome
The evaluation generated 30 written data requests across regulatory, legal, validation, and surveillance domains; three prepared vendor meetings, including a live review of preselected cases; a consecutive-case output log of 400 examinations, including 30 detailed audits; two operating-point analyses; one version-to-version comparison; and three contract, indemnification, and insurance reviews. Estimated effort exceeded 40 radiologist-hours, all outside clinical hours and any research protocol. No work product was contractually scoped, compensated, or attributed, and no site-level performance data was returned. The vendor's written surveillance description stated that site-level false-positive and false-negative rates are not returned to practices by design; instead, surveillance data is aggregated into periodic safety reports for notified bodies and auditors (Figure 2). A successor version later shifted output distribution in the direction previously reported by the site, but no communication identified the site's observations as an input.

Conclusion
The economics are inverted: the party best positioned to observe real-world performance bears the full cost of generating evidence while receiving none of the resulting surveillance data. This persists because post-market evaluation labor remains invisible, unrecognized, and unbudgeted.

Statement of Impact
Evaluation labor should be incorporated into procurement as a defined scope of work with measurable deliverables, compensation or fee offsets, and attribution when site-generated data informs product improvements. Quantifying this hidden labor is a prerequisite for meaningful negotiation and for converting

goodwill into durable surveillance infrastructure.

Evaluation Work Product	Volume	Scoped	Paid	Research	Data Returned	Attribution
Vendor data requests *Regulatory, legal, validation, and surveillance domains*	**30 requests**	No	No	No	No	No
Prepared vendor meetings *Live case review with pre-selected cases prior to the meeting*	**3 meetings**	No	No	No	No	No
Consecutive-case output log *Captured AI output from consecutive clinical examinations during deployment*	**400 examinations**	No	No	No	No	No
Per-case audit Flag status, marker tier, and processing-failure log, case by case	**30 examinations**	No	No	No	No	No
Operating-point analysis *Observed flag rate tested against labeled sensitivity and specificity*	**2 analyses**	No	No	No	No	No
Version comparison Output distribution of deployed vs. successor version on the same cases	**1 analysis**	No	No	No	No	No
Contract review Liability cap, indemnification, and policy jurisdiction	**3 reviews**	No	No	No	No	No
Estimated evaluation effort: >40 radiologist-hours over 5 months *(outside clinical hours)*		**None**	**None**	**None**	**None**	**None**

Figure 1. Ledger of evaluation work products generated during a five-month site-side assessment of an FDA-cleared prostate MRI CAD tool. Each work product was classified according to five attributes: contractual scope, compensation, research-protocol coverage, return of performance data to the site, and attribution in vendor-facing outputs. No work product met any of these criteria.

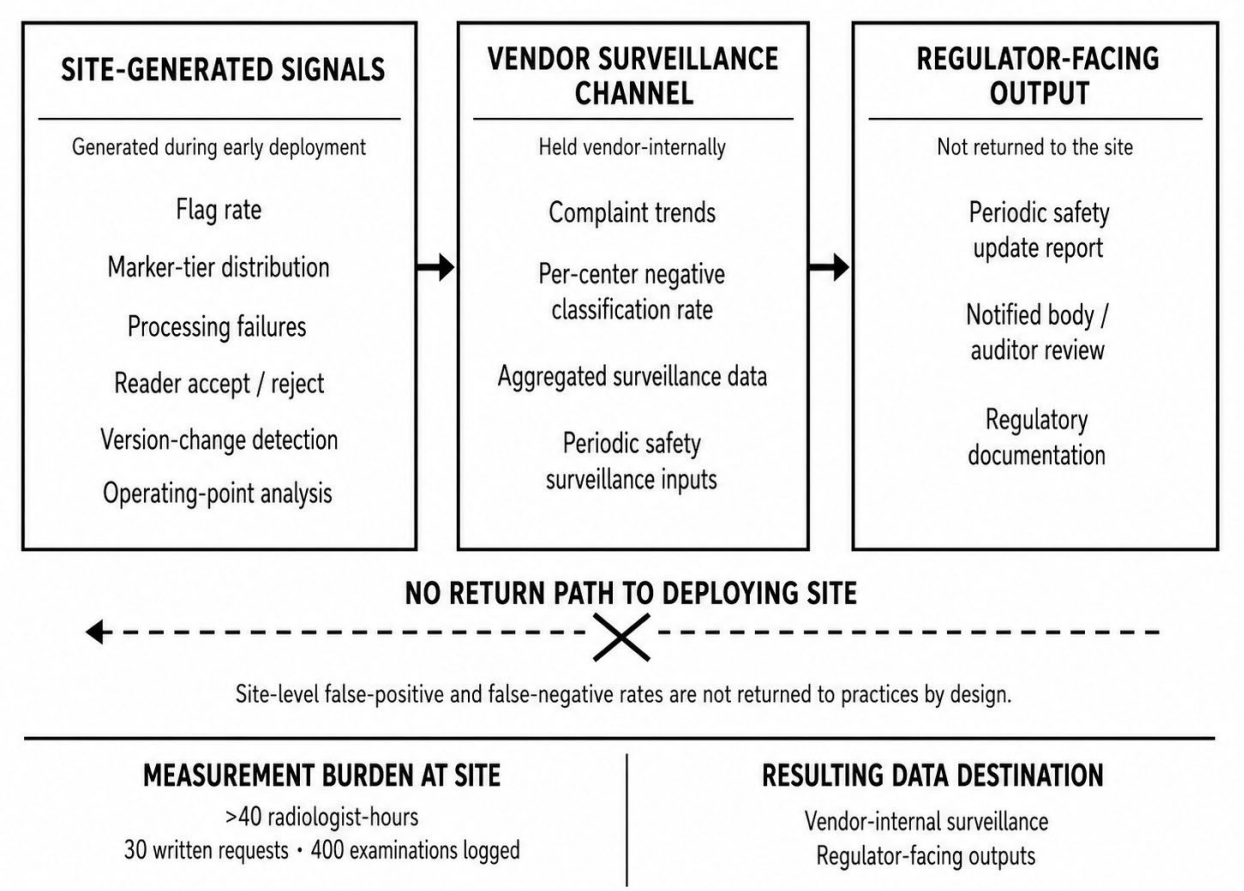

Figure 2. Flow of site-generated performance data during post-market evaluation. Performance signals generated at the deploying site enter vendor-internal surveillance and may contribute to regulator-facing outputs, but site-level false-positive and false-negative rates are not returned to the practice. The site therefore bears the measurement burden while the resulting data remain vendor-internal or regulator-facing

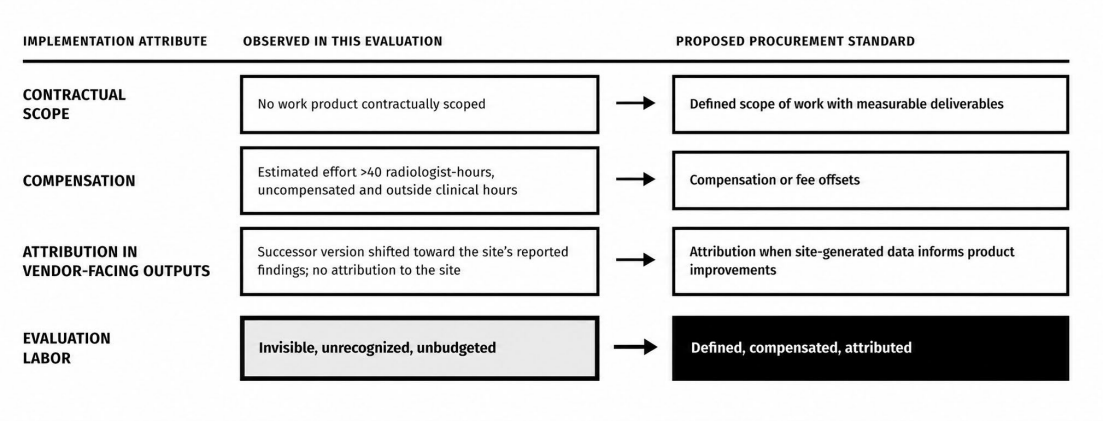

Figure 3. Proposed procurement standard for site-side evaluation labor. Three implementation attributes unmet during this evaluation are paired with the standard proposed for future purchase agreements.

Keywords

post-market surveillance; clinical AI implementation; vendor evaluation; radiologist workload; procurement governance