Assessing the Long-Term Stability of Reference Dosimetry Systems in a Physics-Consulting Intercomparison Network, 2002 to 2026

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

Ethan Cohen[1]; Hunter Tillery[1]
[1] Northwest Medical Physics Center, Lynnwood, WA, USA

ABSTRACT

Purpose
To characterize the long-term stability and reproducibility of reference-class dosimetry systems through a sustained field-kit intercomparison, and to draw QA inferences for consulting practices serving many clinics from a shared instrument pool.

Methods
We analyzed 46 semiannual events conducted by NMPC from September 2002 through March 2026. Each center's chamber–electrometer kit is irradiated in a shared 6 MV host-linac beam and referenced to a NIST-traceable Local Standard; the figure of merit is the output ratio (field kit / Local Standard). HDR well chambers are separately referenced to a Local-Standard well chamber and to the planning-predicted air-kerma strength. Results tracked by chamber serial number across 1,255 kit measurements.

Results
Output ratios clustered near unity (grand mean 0.998, SD 0.008; 99.8% within 2%). Per-chamber reproducibility was high (SD ≤ 0.007 for nearly all serials tracked ≥ 3 events). The program detected discrete equipment faults—an uncalibrated chamber, an HDR afterloader source-strength offset, and an out-of-tolerance electrometer—that would not surface in routine in-clinic QA.

Conclusions
Well-maintained reference systems hold sub-percent stability over more than two decades. Semiannual intercomparison resolves drift from noise and detects faults across the deployed measurement chain, complementing biennial ADCL recalibration.

CONTACT

Ethan Cohen
Northwest Medical Physics Center
ECohen@nmpc.org

INTRODUCTION

External dosimetry audits underpin confidence in reference beam calibration. Mailed TLD/OSLD audits verify an institution's beam output using a transfer detector,[1,2] but they assess the *delivered beam* rather than the dosimetry instrument itself and carry transport/readout uncertainty of ~1.3–1.6%.[2]

NMPC has operated a complementary audit since 2002: a **field-kit-to-Local-Standard intercomparison** in which each center's own chamber–electrometer kit is irradiated in a shared host beam against a co-located reference. This design audits the *deployed measurement instrument* (chamber, electrometer, thermometer, barometer) under common-beam conditions, isolating instrument behavior from linac output.

Sustained over 46 semiannual events and 1,255 measurements, the record permits questions single audits cannot address:

- How stable are reference chambers over a working lifetime?
- How reproducible is the measurement chain?
- What fault modes does periodic intercomparison catch?

METHODS AND MATERIALS

Design
Each event: field kits irradiated sequentially in shared host-linac photon beam (6 MV, 10×10 cm, acrylic phantom, chamber at 100 cm SSD). Output ratio = field kit reading / Local Standard reading.

Local Standard
PTW 0.6 cc Farmer-type chamber (S/N 91) with CNMC 206 electrometer, in continuous service since 2002 with NIST-traceable ADCL calibration. Local Standard measured at start and end of each linac run and averaged.

Formalism
Early events used TG-21 formalism;[5] cross-check tables first appear 2003; full transition to TG-51 $N_{D,w}$ output-ratio basis by 2008.[6,7]

Environmental
Each kit's thermometer and barometer compared to Local-Standard instruments. Field-vs-reference C_{TP} deviations on order of 0.1–0.3%.

HDR
From April 2012: source driven sequentially into each well chamber via transfer tube; referenced to Local-Standard well chamber (SI HDR 1000 Plus) and planning-predicted air-kerma strength.[8,9]

Acceptance
Field kit accepted when output ratio within **±2%** of Local Standard. HDR acceptance: ~2% vs Local Standard, ~5% vs source certificate.

RESULTS

1,255	46	99.8%
Total Kit Measurements	Semiannual Events	Within ±2%

Aggregate Stability
Grand-mean output ratio: **0.998** (SD 0.008). Per-event means: 0.991–1.005; within-event SD: 0.002–0.011. No secular trend in cohort mean.

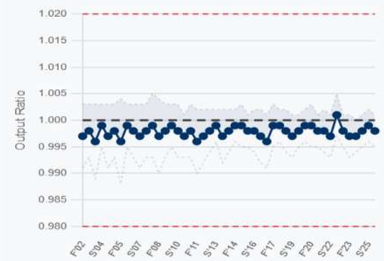

Figure 1. Per-event mean output ratio (±1 SD) over 46 semiannual intercomparison events, 2002–2026. Dashed lines indicate ±2% action levels.

Per-Chamber Reproducibility (Serial-Tracked)
Of 54 chambers measured in ≥ 3 events: per-chamber SD ≤ 0.007 for nearly all; net drift (last − first) within ±0.02 for every chamber and ±0.01 for most. Chambers reassigned between centers showed no step change in ratio.

Chamber (Model, S/N)	N	Mean	SD	Range	Net Drift
PTW 30001, 488	46	0.996	0.006	0.023	+0.010
PTW 30013, 1969	41	0.997	0.005	0.034	−0.014
PTW 23333, 1936	39	0.998	0.007	0.030	−0.003
PTW 30001, 1891	37	0.998	0.004	0.021	−0.003
PTW 30013, 790	25	0.996	0.007	0.035	+0.007
PTW 30013, 2840	27	0.999	0.004	0.019	−0.004

Table 1. Representative chambers tracked by serial number. S/N 790, 1969, 2840 reassigned between centers with no step change.

Faults Detected:
① Uncalibrated chamber (Oct 2008, S/N 3111, ratio 1.216) — flagged, excluded
② HDR afterloader source-strength offset (2014–15): all chambers read 2.4–2.7% low vs planning while agreeing with each other — localized to afterloader, not kits
③ Out-of-tolerance electrometer (Spring 2025, ratio 1.022): confirmed by re-measurement with LS electrometer (0.995) — returned to manufacturer

DISCUSSION

Two decades of common-beam intercomparison show that well-maintained reference chambers and electrometers hold **sub-percent stability** over their working lifetimes. The dominant variation is within-event measurement spread rather than long-term drift.

The ADCL stability criterion is an old-vs-new $N_{D,w}$ ratio SD below ~0.3% for repeat calibrations of the same chamber.[10] The NMPC in-beam per-chamber SD (≤ 0.7%) is larger because it folds in host-beam and setup variation, yet confirms comparable stability.

The **serial-number view** disentangles instrument behavior from clinic identity: a chamber's ratio history is continuous across site reassignments, so per-kit longitudinal tracking—not per-site averaging—is the appropriate lens.

Key insight: The HDR afterloader offset illustrates the program's distinctive value. Its common-mode signature across every well chamber—absent in the chamber-to-chamber comparison—localized the error to the treatment unit's stated source strength rather than to any dosimetry kit.[9]

An ADCL calibrates chamber and electrometer together, but only periodically (~biennial) and apart from the clinic's thermometer, barometer, and treatment unit. Omitting periodic stability checks can add 0.3–0.5% uncertainty.[7] Intercomparison supplies verification *frequently* and *in-beam* against a peer cohort.

The semiannual cadence is integral: distinguishing progressive drift from single-event noise requires repeated sampling. Consistent with this, 42% of confirmed calibration errors in a multi-network audit were identified during ongoing periodic monitoring rather than at first audit.[1]

CONCLUSIONS

- Reference dosimetry systems achieve and maintain sub-percent stability over > 2 decades when supported by rigorous QA
- The program's principal contribution is detecting equipment faults across the measurement chain—and a treatment-unit source-strength error—that routine institutional QA need not reveal
- Semiannual intercomparison provides temporal resolution to separate drift from noise, complementing biennial ADCL recalibration
- Per-chamber serial-number tracking, not per-site averaging, is the appropriate framework for consulting-network intercomparison

REFERENCES

1. Kry SF et al. Remote beam output audits: a global assessment. Phys Imaging Radiat Oncol. 2018;7:39-44.
2. Alvarez P et al. TLD and OSLD dosimetry systems for remote audits. Radiat Meas. 2017;106:412-415.
3. Gershkevitsh E et al. Dosimetric inter-institutional comparison. Acta Oncol. 2014;53(5):628-636.
4. Beveridge S et al. International film dosimetry intercomparison. Med Phys. 2024;51(12):9071-9087.
5. TG-21. A protocol for absorbed dose from high-energy beams. Med Phys. 1983;10(6):741-771.
6. Almond PR et al. AAPM's TG-51 protocol. Med Phys. 1999;26(9):1847-1870.
7. McEwen M et al. TG-51 addendum. Med Phys. 2014;41(4):041501.
8. Rivard MJ et al. Update of TG-43 Report. Med Phys. 2004;31(3):633-674.
9. Rasmussen BE et al. Comparison of air-kerma strength for HDR ^{192}Ir. Med Phys. 2011;38(12):6721.
10. Muir BR. N_D,w measured at ADCLs: stability. Med Phys. 2015;42(4):1546-1554.