# Use Case: Applying the Digital Health Canada Clinical AI Evaluation Toolkit in Practice

## Context

This example illustrates how the Clinical AI Evaluation Toolkit can be applied to a real-world deployment of an ambient clinical documentation tool within a large Canadian academic healthcare organization.

The tool was introduced to support clinical documentation by generating draft notes during patient encounters. The evaluation below reflects findings approximately 9 weeks post-launch and follows the toolkit's Minimum Viable Evaluation (MVE) pathway, with selected additional indicators and a directional ROI assessment.

## Prepared for

Project champion and team

## Purpose

The toolkit applied to the Ambient clinical documentation tool project at this point in time, producing the toolkit's actual outputs — Readiness Checklist, MVE pre/post comparison, Expanded Pathway KPI status, and directional ROI calculations — populated with current data and labeled benchmark estimates where institutional values are pending.

## Status

Interim evaluation output. Some indicators are post-deployment and current; others are pending August 2026 follow-up. ROI calculations use benchmark estimates clearly flagged in the inputs table.

## Executive summary

The toolkit's evaluation pathway, applied to the ambient clinical documentation tool project at 9 weeks post-launch, supports the following findings:

- **Readiness:** Pre-deployment readiness was met at minimum levels across all seven dimensions; three are met without qualifications, three are partial pending future evaluation activities, and one is met with a noted qualification.
- **Adoption:** 44% of the pilot cohort reached meaningful adoption (≥10% of encounters) at 9 weeks post-launch, robust to threshold sensitivity. 66% used the tool at least once. Pulse-check feedback from 40% of respondents reported improved eye contact and reduced screen time, alongside concerns about psychiatric content omission and AI anchoring of clinical judgment.
- **Outcomes:** Pre-deployment trends (rising documentation time, rising after-hours work, stable burnout prevalence) indicate continuing burden; full post-deployment data pending August 2026.
- **Directional ROI:** At benchmark inputs (CA$350 fully-loaded hourly rate; 30 encounters/physician/week; 44 clinical weeks/year), the directional ROI signal at 9 weeks post-launch ranges from CA$716,000 (low scenario, 3 min saved/encounter, current 31-physician cohort) to CA$2,387,000 (high scenario, 10 min saved). Conservative variants using ≥20% and ≥30% adoption thresholds yield CA$416,000 and CA$254,000 respectively.
- **Decision position:** Continue with conditions, pending resolution of identified quality and workflow risks.

The primary risk at this stage is documentation quality and associated editing burden, which may offset time savings if not addressed.

# 1. Implementation Readiness Checklist (completed)

The toolkit's Readiness Checklist documents pre-deployment conditions. Stricter grading is applied: dimensions without active qualifications are Met; dimensions where data are not yet collected or full execution pending are marked Partial.

| Readiness dimension | Status | Evidence |
|---|---|---|
| **Workflow fit** | **Met** | Active EHR encounter required before generation. Standard workflows defined for in-person and virtual visits across 7 subspecialties. Encounter-association troubleshooting included in training. |
| **Data basics** | **Partial** | Vendor-provided enterprise-integrated platform within institutional governance. One-month vendor analytics interruption (Feb 17–Mar 18, 2026) means utilization data from early rollout is incomplete. |
| **Staff comfort** | **Met** | One-to-one onboarding, superuser support, tip sheets from day 1. Pulse check 27/67 (40%) initial response. Ad-hoc training-period testing addressed physician-raised issues. |
| **Time and capacity** | **Met** | 14-month governance preparation. Steering committee, PIA, TRA, environmental scan, plain-language specialist review all completed pre-launch. |
| **Context factors** | **Met** | Five-year longitudinal EHR-burden program at the academic healthcare organization. provides quantified pre-deployment baseline. QPER protocol 169. |
| **Equity and experience** | **Partial** | Patient experience instrument designed with disaggregated race/ethnicity, gender (non-binary, Two-Spirit), first-language items. Equity outcomes data not yet collected. |
| **Evaluation feasibility** | **Partial** | Mixed-methods QI design informed by TEHAI framework. REDCap surveys at pre, 2-month, 6-month. PDQI-9 chart sampling planned. Full execution pending August 2026. |

### Readiness summary

- 3 of 7 dimensions are Met without qualification (Workflow fit, Time and capacity, Context factors)
- 1 of 7 is Met with a noted qualification (Staff comfort: modest pulse-check response)
- 3 of 7 are Partial pending future evaluation events (Data basics: vendor analytics gap; Equity: outcomes data not yet collected; Evaluation feasibility: full execution pending)

This grading is conservative; a permissive reading against toolkit minimums would mark all seven Met. The conservative grading is more useful for honest organizational tracking.

## 2. MVE Pathway: pre/post comparison

The toolkit's MVE pathway calls for indicators measured at baseline and post-deployment. The ambient clinical documentation tool project has unusual baseline strength: a five-year longitudinal EHR-burden program at the academic healthcare organization provides pre-deployment data on three of the four physician-experience indicators. Adoption and clinician feedback are post-deployment only.

| Indicator | Pre baseline | Mid baseline | Post (current) | Interpretation |
|---|---|---|---|---|
| **Burnout prevalence (Mini-Z)** | 25.6% (2019) | 26% (2024) | Pending Aug 2026 follow-up | Stable; full post-deployment Mini-Z scheduled at 6-month mark |
| **Documentation time per visit** | 12.2 min (2020) | 13.2 min (2024) | Pending post-deployment monitoring | Pre-deployment trend rising; post-deployment to determine inflection |
| **After-hours EHR use** | 19.7% (2020) | 21.0% (2024) | Pending post-deployment monitoring | Pre-deployment trend rising; post-deployment direction unknown |
| **EHR use as burnout contributor** | ~75% of burned-out physicians (2019) | 52% of burned-out physicians (2024) | Pending Aug 2026 Mini-Z follow-up | Subjective improvement despite objective rise; ambient clinical documentation tool motivated by closing this perception-reality gap |
| **Tool adoption (≥10% encounters)** | n/a (pre-deployment) | n/a (pre-deployment) | 31/70 (44%) at 9 weeks post-launch | Threshold sensitivity: 18 (26%) at ≥20%; 11 (16%) at ≥30% |
| **Tool adoption (used at least once)** | n/a | n/a | 46/70 (66%) | Modestly undercounted by 1-month vendor analytics gap |
| **Pulse-check clinician feedback** | n/a | n/a | 27/67 (40%) responding | Improved eye contact, reduced screen time; concerns re psychiatric content omission and anchoring risk |

### MVE interpretation

The pre-deployment trend across four physician-experience indicators is mixed: subjective burden attribution to EHR improved (75%→52%) while objective documentation time and after-hours work both rose. This perception-reality gap was the explicit motivation for the ambient clinical documentation tool deployment. Post-deployment data through August 2026 will determine whether the deployment closes the gap on the objective indicators while sustaining or improving subjective measures.

Adoption findings at 9 weeks are robust: 44% reach meaningful (≥10%) adoption; the figure holds at lower levels under stricter thresholds (26% at ≥20%, 16% at ≥30%), suggesting adoption findings are not driven by the choice of cutoff. The 1-month vendor analytics gap likely modestly undercounts true at-least-once usage but does not change directional findings.

## 3. Expanded Pathway: 4 KPIs and current values

The toolkit's Expanded Pathway permits additional KPIs from the categorized library. Four (4) KPIs were selected to fill gaps in the MVE indicator set. Current values are populated where available.

| KPI | Domain | Current value | Notes |
|---|---|---|---|
| **Frequency of manual edits required** | Quality | Not yet captured numerically | Qualitative pulse-check signals (omissions of psychiatric content, compression of clinically significant detail) suggest substantial editing required; planned PDQI-9 chart sampling will quantify |
| **Reported equity concerns** | Equity | Performance gaps in identity-sensitive rendering identified | Misgendering of gender-diverse patients; inaccurate rendering of names from non-Western backgrounds; consent challenges in psychosis-spectrum populations |
| **Training completion** | Experience | 70/70 pilot physicians completed onboarding | 100%; one-to-one onboarding model used |
| **Model latency** | Time | Not yet captured | Vendor performance dashboard available; metric not yet pulled into the organization's evaluation reporting |

Of the four added KPIs, training completion (100%) and reported equity concerns (specific performance gaps named) have current values. Frequency of manual edits and model latency are not yet captured at the metric level. The frequency-of-edits gap is the most consequential to close before the 6-month evaluation, since it converts qualitative draft-quality concerns into a numeric indicator that PDQI-9 chart sampling can validate.

These indicators highlight important areas requiring further measurement before scaling decisions.

## 4. Directional ROI calculation (early estimate)

The toolkit's Minimum ROI Calculator (directional tier) takes inputs that produce a 12-month time-value figure. The output is explicitly directional rather than audited. Inputs marked BENCHMARK ESTIMATE use industry appropriate benchmarks; institutional values should replace these when available.

These estimates are directional and contingent on confirmed time savings and documentation quality at follow-up.

### Inputs

| Input | Value | Source / status |
|---|---|---|
| **Perspective** | Institutional (the academic healthcare organization) | Confirmed |
| **Comparator** | Standard manual documentation pre-deployment | Confirmed |
| **Time horizon** | 12 months from go-live | Confirmed |
| **Adopting cohort (≥10% threshold)** | 31 physicians | Confirmed |
| **Adopting cohort (≥20% threshold)** | 18 physicians | Confirmed (sensitivity analysis) |
| **Adopting cohort (≥30% threshold)** | 11 physicians | Confirmed (sensitivity analysis) |
| **Encounters per physician per week** | 30 | BENCHMARK ESTIMATE (typical Canadian outpatient specialist physician: 6 encounters/ day × 5 days; varies by subspecialty and FTE) |
| **Clinical weeks per year** | 44 | BENCHMARK ESTIMATE (52 weeks − 8 weeks vacation/conference/leave) |
| **Time saved per encounter — Low** | 3 minutes | BENCHMARK ESTIMATE (conservative; consistent with low-end ambient clinical documentation tool literature) |
| **Time saved per encounter — Medium** | 6 minutes | BENCHMARK ESTIMATE (consistent with mid-range ambient clinical documentation tool primary care reports, e.g., Stults 2025, Ma 2025) |
| **Time saved per encounter — High** | 10 minutes | BENCHMARK ESTIMATE (high-end; reported in some primary-care studies but unlikely to fully transfer to psychiatry) |
| **Fully-loaded hourly cost (specialist physician)** | $350 CAD | BENCHMARK ESTIMATE (Canadian academic-hospital specialist physician FLR; range $300–$420 in published HR benchmarks) |
| **Implementation cost (14-month preparation)** | Not yet itemized | TO BE PULLED FROM THE ACADEMIC HEALTHCARE ORGANIZATION INTERNAL RECORDS |

**Scenario outputs (directional)**

Five scenarios are computed from the inputs above. The first three vary time saved per encounter at the current 31-physician adopting cohort (≥10% threshold). The last two model conservative variants using stricter adoption thresholds (≥20% and ≥30%).

| Scenario | Adopting cohort | Encounters / yr | Time returned | Directional value (CAD) |
|---|---|---|---|---|
| **Low (3 min/ encounter)** | 31 (≥10%) | 40,920 / yr | 2,046 hours | **~$716,000** |
| **Medium (6 min/ encounter)** | 31 (≥10%) | 40,920 / yr | 4,092 hours | **~$1,432,000** |
| **High (10 min/ encounter)** | 31 (≥10%) | 40,920 / yr | 6,820 hours | **~$2,387,000** |
| **Conservative (3 min, ≥20% threshold)** | 18 (≥20%) | 23,760 / yr | 1,188 hours | **~$416,000** |
| **Conservative (3 min, ≥30% threshold)** | 11 (≥30%) | 14,520 / yr | 726 hours | **~$254,000** |

### ROI interpretation

- **Direction:** All scenarios produce positive directional time-value at 12 months relative to the cohort's clinical effort.
- **Range:** The three-scenario range at the current cohort spans roughly CA$716K to CA$2.4M. Most ambient clinical documentation tool literature in primary care reports time savings in the 3–6 minute range; the high-end (10 min) figure should be treated as an upper bound that may not transfer to psychiatry given the manuscript's specialty-specific findings.
- **Sensitivity:** At ≥20% threshold (18 physicians) and conservative time saved (3 min), directional value remains substantial (CA$416K). Even the most conservative scenario (≥30% threshold, 3 min) yields ~CA$254K.
- **Implementation cost:** Implementation cost is not yet itemized and is not deducted from the figures above. A 14-month preparation cost in the low six figures (steering committee time, PIA, TRA, environmental scan, plain-language specialist, training) is plausible. Net direction would remain positive across all scenarios but warrants confirmation.
- **Non-monetized benefits:** Quality, equity, and clinician experience benefits are not monetized. The directional ROI captures only physician time return; the full case for the deployment includes non-monetized clinical and equity benefits.

## 5. Findings narrative

### What the toolkit application reveals at this point

Applied at 9 weeks post-launch, the toolkit produces three findings:

- **Strong readiness foundation:** The Ambient clinical documentation tool project entered deployment with unusually strong pre-deployment readiness and a quantified longitudinal baseline that few comparable Canadian implementations have available. Three of seven readiness dimensions are unqualified Met, including Context factors — the dimension where most Canadian implementations would not match the academic healthcare organization's baseline strength.
- **Early adoption signal positive but specialty caveats real:** 9-week adoption (44% at ≥10% threshold; 66% used at least once) is robust to threshold sensitivity. Pulse-check signals are positive on physician experience but flag specialty-specific concerns about psychiatric content omission and AI anchoring of clinical judgment. These signals motivate the planned PDQI-9 chart sampling, which will quantify draft quality systematically.
- **Directional ROI is positive across scenarios:** Even at conservative inputs, directional ROI is positive at 12 months. The strongest directional signal is sensitivity-stable: at ≥20% threshold and conservative 3-minute time savings, directional value remains roughly CA$416K. The full case for the deployment also includes quality, equity, and clinician experience benefits not captured in the directional ROI.

### What the toolkit application does not yet show

- Post-deployment Mini-Z burnout, documentation time, after-hours EHR use — pending August 2026
- PDQI-9 draft quality measurement — chart sampling pending
- Equity-disaggregated patient experience outcomes — survey administration ongoing
- Frequency of manual edits at the numeric level — measurement method to be confirmed (vendor analytics or sampled chart audit)
- Implementation cost itemization — to be pulled from the academic healthcare organization internal records before final ROI
- Whether the trust-first implementation approach produced better outcomes than a standard deployment would have — not measurable from current data; comparative benchmarking with other Canadian implementations could partially address

## 6. Guardrails (toolkit)

Per the toolkit's stated guardrails:

- **Evaluation is not validation:** All outputs above support structured learning, not regulatory or formal validation. The academic healthcare organization's QPER protocol 169 governs ethics oversight separately.
- **Directional ROI:** ROI scenarios are directional, not audited. Benchmark estimates flagged where used.
- **Minimum means minimum:** MVE pathway is sufficient for the early evaluation reported here. Expanded Pathway components are additive.
- **Ongoing oversight:** Findings will be revisited at the 6-month QPER evaluation and quarterly thereafter.
- **Responsible AI:** Safety, equity, transparency, sustainability, stewardship considerations are integrated through the patient-experience instrument design, the Patient and Family Education team review, and the planned focus group with lived-experience partners.

## 7. Action items emerging from this application

- Pull the academic healthcare organization HR fully-loaded hourly rate for specialist physicians; replace benchmark estimate in ROI inputs
- Pull encounter volume actuals (the academic healthcare organization scheduling/billing) for the 70-physician cohort; replace benchmark estimate
- Itemize 14-month implementation cost from internal records; complete net ROI calculation
- Confirm with vendor whether analytics dashboard provides draft-edit retention metric; if not, plan sampled chart audit (50 notes per subspecialty) as part of 6-month QPER evaluation
- Add reported-equity-concerns item to next pulse-check distribution
- Schedule quarterly ongoing-oversight reviews post-August 2026 with QPER governance

## 8. Decision Guidance

Decision position: Continue with conditions

Adoption: Early adoption is meaningful but not yet widespread, indicating initial traction without full normalization across users

- Risk signals: Identified concerns related to documentation quality, editing burden, and equity require further measurement and mitigation
- Value signal: Directional ROI is positive across scenarios, but dependent on assumptions and not yet validated with observed post-deployment outcomes

Implication: Continue implementation with targeted focus on:

- quantifying and reducing documentation editing burden
- validating actual time savings in practice
- addressing identified content quality and equity concerns

Reassessment is recommended at the 6-month evaluation point before broader scale decisions.

## What This Does Not Yet Show

At this stage, several key indicators remain in progress:

- Post-deployment changes in documentation time and after-hours work
- Quantitative assessment of documentation quality
- Equity-disaggregated patient experience outcomes
- Confirmed time savings and full ROI inputs
- Implementation cost and net value

## Key Takeaway

Early signals are promising, with meaningful adoption and positive directional value.

However, important risks related to documentation quality, workflow impact, and equity must be addressed before scaling.

This example demonstrates how the toolkit supports structured, evidence-informed decision-making in early-stage AI deployment, without requiring full or perfect data.