QUALITY INTELLIGENCE INSTITUTE

DEFINING THE FUTURE OF HEALTHCARE AI QUALITY.

QUALITY INTELLIGENCE INSTITUTE

The State of AI in Medicare Advantage Quality

What Medicare Advantage leaders should scale, pilot, test—and refuse to overclaim

QII Research Briefing | PUB-RB-001 | August 2026

Executive Summary

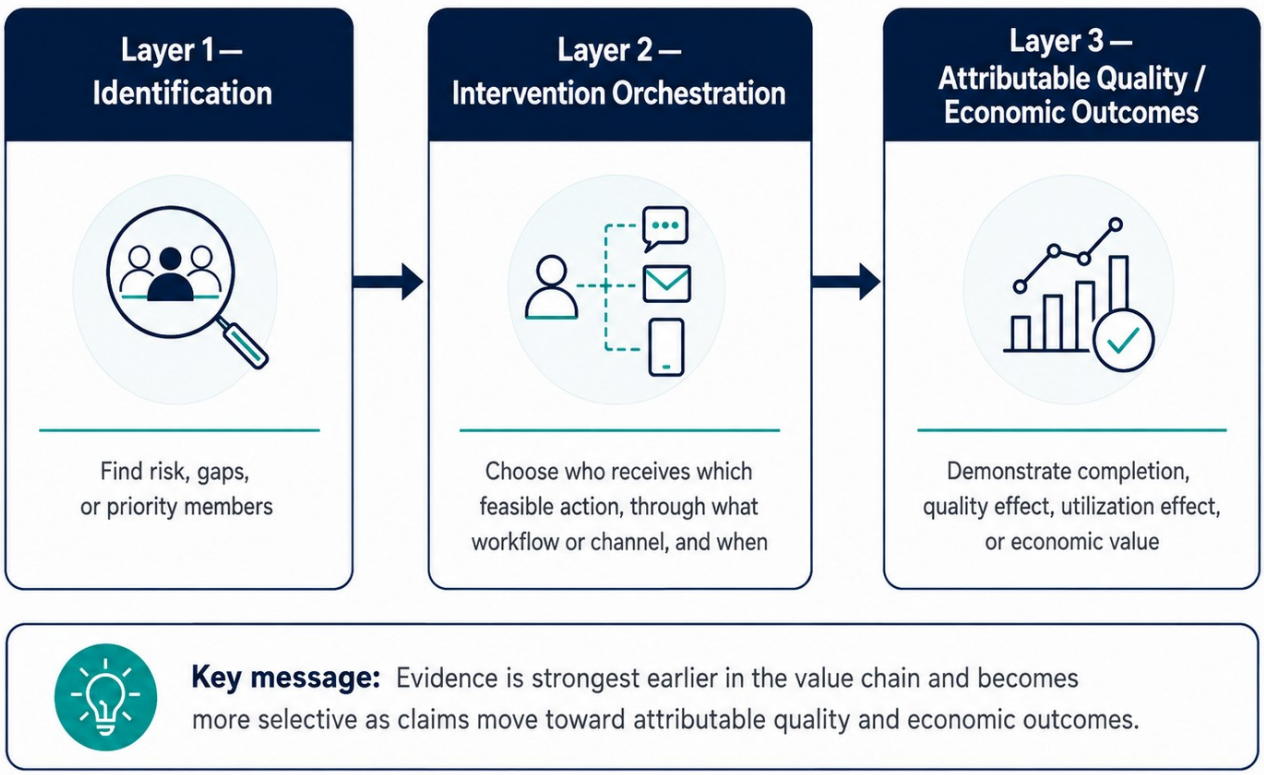
Artificial intelligence is beginning to produce measurable quality value—not because algorithms act alone, but because they are being embedded in intervention systems that decide who receives attention, how scarce care-management capacity is allocated, and how quickly quality feedback reaches clinicians. Across the evidence reviewed by QII, the strongest results now extend beyond prediction to actual screening completion, medication adherence, selected utilization outcomes, gross medical-cost effects and improvement in a CMS process measure.[6,19–25]

The evidence becomes less mature when organizations claim that AI itself caused contract-level HEDIS or Star improvement, or when gross savings are presented as net AI ROI. For Medicare Advantage leaders, the practical question is therefore not simply whether an AI model is accurate. It is whether the organization has a feasible intervention to act on the output, can measure whether that intervention was completed, can attribute the resulting quality effect at the level claimed, and can demonstrate value after the full cost of making the human-AI system work.

QII Three-Layer Model of AI Value
Layer 1 — Identification: find risk, gaps or members who should be prioritized. Layer 2 — Intervention Orchestration: determine who receives which feasible action, through which workflow or channel, and when. Layer 3 — Attributable Quality / Economic Outcomes: demonstrate completed care, measure improvement, utilization effects, contract-level performance or net financial value. The reviewed evidence is strongest in Layers 1 and 2 and increasingly selective in Layer 3. [5,6,19-25]

Figure 1. QII Three-Layer Model of AI Value

Illustrative conceptual model for Medicare Advantage quality AI value



Five conclusions for MA executives

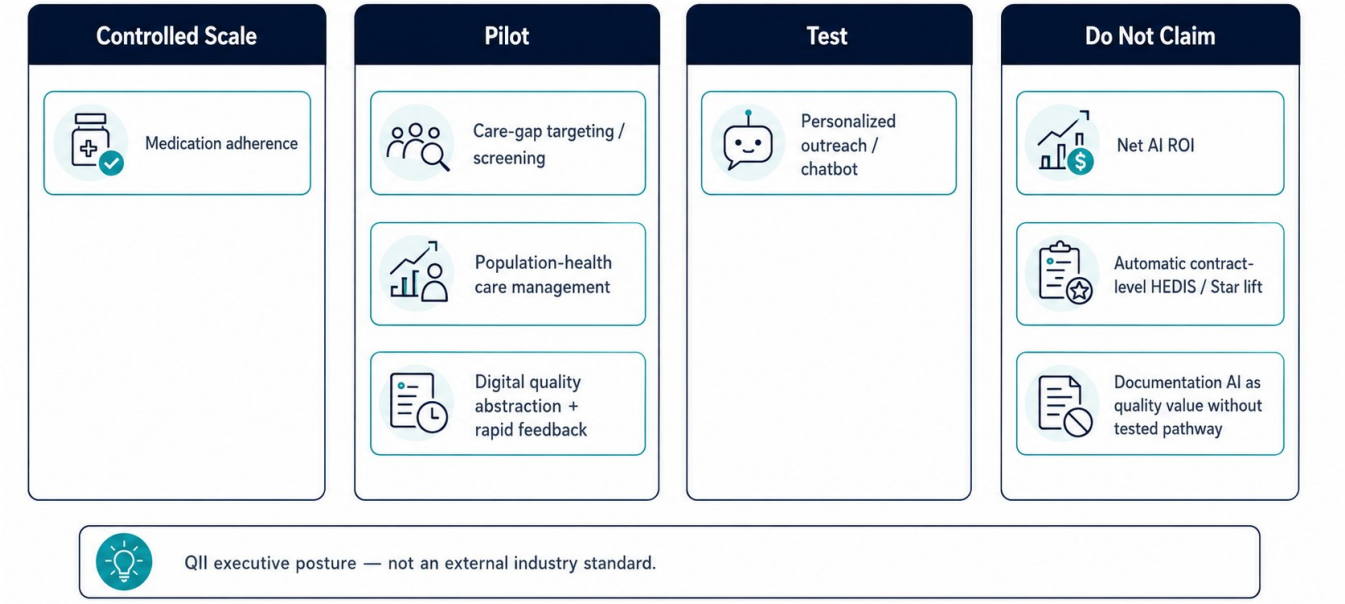
Executive conclusion	What it means
1. Evidence now reaches beyond prediction	AI/ML-guided programs have demonstrated actual screening completion, selected adherence improvement, utilization effects, gross medical-cost effects and improvement in a CMS process measure.[19–25]
2. The positive evidence is usually program-level, not algorithm-only	Human outreach, pharmacists, care managers and rapid feedback are part of the effective intervention system. Attribution to the AI component alone remains difficult in many studies.[19,20,22,23,25]
3. Personalization is not automatically effective	Adaptive personalization improved adherence in REINFORCE, but a much larger pragmatic trial found no sustained 12-month benefit from reminder/nudge/chatbot strategies after correction for multiple comparisons.[6,21]
4. MA-specific economic effects exist, but net AI ROI is a higher bar	Predictive-analytics-guided outreach in MA was associated with lower utilization and $716 lower total medical cost per month during the first six months, but full technology and operating costs were not incorporated into an AI-specific ROI calculation.[22]
5. Contract-level HEDIS/Star attribution remains a decision boundary	Medicare-relevant adherence and reported Star improvement exist in a bundled AI-supported pharmacist program, and actual care-gap completion is now demonstrated elsewhere, but AI-alone causal effects on contract-level HEDIS/Stars remain limited.[19,20]

What the evidence is strong enough to do now

Use case	QII executive posture	Evidence position	Decision requirement
Medication adherence	**CONTROLLED SCALE / PILOT**	Measure-level and reported Star improvement exist in an AI-supported pharmacist bundle; adaptive RCT evidence is positive in one setting.[6,19,21]	Require program-level comparator, PDC/Star endpoint, pharmacist/workflow attribution and sustainability measurement.
Care-gap targeting / screening	**PILOT / SCALE WHERE WORKFLOW EXISTS**	Direct ML-guided screening completion demonstrated.[20]	Require actual completion endpoint and intervention-capacity plan; do not accept model lift as closure.
Population-health care management	**TARGETED PILOT / SCALE**	MA payer and contemporary AI-guided care-management outcome evidence exist.[22,23]	Define target population, human intervention, utilization endpoint and comparator; separate algorithm and care-management contribution where possible.
Digital quality abstraction + rapid feedback	**PILOT**	Randomized CMS process-measure improvement demonstrated outside HEDIS/MA.[25]	Require accuracy, timeliness, human review and downstream measure-effect testing in the intended workflow.
Personalized outreach / chatbot	**TEST—DO NOT ASSUME**	Evidence is heterogeneous; a large pragmatic trial found no sustained benefit.[6,8,21]	Require a completion endpoint and escalation strategy beyond low-intensity digital messaging.
Ambient documentation AI	**BUY FOR WORKFLOW UNLESS QUALITY PATHWAY IS TESTED**	Randomized workflow benefits exist; HEDIS/Star effects do not.[13,14]	Treat as documentation/workforce ROI unless the pathway to usable quality data or clinical action is separately validated.
AI ROI claim	**DO NOT ACCEPT WITHOUT FULL COST STACK**	Gross medical-cost effects exist; net AI ROI remains underdeveloped.[17,22]	Require attributable benefits plus technology, integration, intervention labor, monitoring, governance and time horizon.

Figure 3. What the Evidence Is Strong Enough to Do Now

QII executive posture by use case



AI Value Realization Gap™

QII uses AI Value Realization Gap™ to describe the distance between evidence that an AI-enabled capability works technically or operationally and evidence that it produces completed interventions, attributable quality outcomes and net economic value. In this briefing, the construct is qualitative and developmental—not a validated numerical benchmark.

1. Why Medicare Advantage Quality Is a Distinct AI Use Case

Medicare Advantage quality sits at the intersection of clinical performance, member experience, operational execution and payment. CMS states that Star Ratings help beneficiaries compare plan quality and are used to determine Quality Bonus Payments and affect rebates. The 2027 Star Ratings measure set continues to include the Part D adherence measures for diabetes medications, hypertension medications (RAS antagonists) and cholesterol medications (statins).[1,26]

That structure makes evidence discipline especially important. An AI-enabled program may improve a screening-completion rate, documentation workflow, hospital process measure or medical spending without necessarily moving a Medicare Advantage contract's HEDIS or Star result. Conversely, an intervention can create operational value even when a contract-level measure effect cannot yet be isolated.

Executive implication

Evaluate every quality-AI investment against the specific endpoint it is expected to influence. If the business case is 'improve Stars,' patient-level or workflow evidence is not enough by itself; the organization needs a defensible pathway to the relevant measure and contract result.

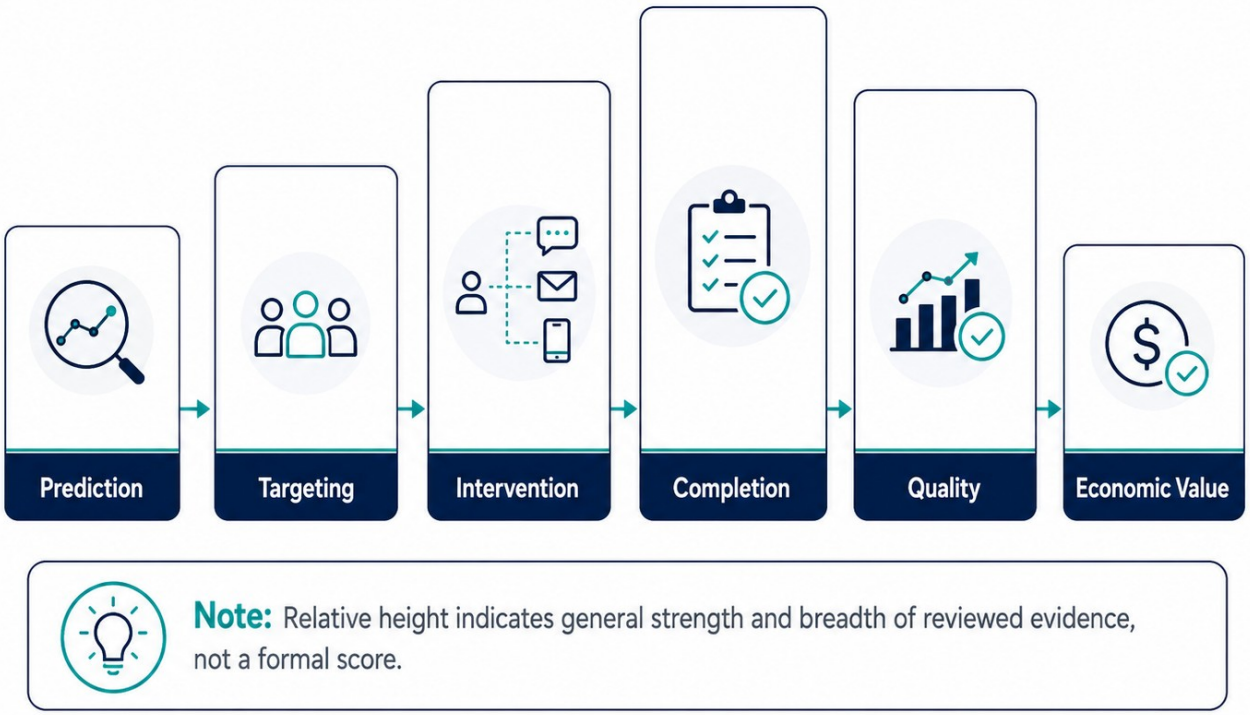
2. The Three-Layer Model: Where AI Value Is Actually Realized

QII organizes the evidence into three layers. The model is derived from the QII Healthcare AI Quality Framework™ and is used here as an analytic decision tool, not an external consensus standard.

Layer	Executive question	Evidence now available
Layer 1 — Identification	Who is at risk, who has a gap, or who should be prioritized?	Strong prediction/prioritization evidence; large payer/population-health implementations; insurer production deployment.[5,24]
Layer 2 — Intervention Orchestration	Who should receive which feasible intervention, through which workflow/channel, and when?	ML-guided screening outreach, adaptive messaging, risk-targeted care management and AI-enabled rapid quality feedback.[6,20,23,25]
Layer 3 — Attributable Outcomes	Did completed care, quality performance, utilization or economics improve—and can the change be attributed at the level claimed?	Selected completion, process-measure, utilization and gross-cost effects exist; contract-level HEDIS/Star attribution and net ROI remain thinner.[19–25]

Figure 2. Evidence Maturity Across the AI Quality Value Chain

Illustrative relative pattern only — not a validated quantitative score



The underlying research designs establish different levels of confidence. Prediction-model reporting, live AI evaluation and intervention trials are distinct evidentiary tasks, reflected in TRIPOD+AI, DECIDE-AI and CONSORT-AI.[2–4] For executive decision-making, the practical rule is simpler: a claim should not jump from one layer to a stronger layer without evidence.

Executive implication

Prioritize use cases where Layer 1 intelligence can be linked to a feasible Layer 2 intervention and a measurable Layer 3 endpoint. A technically impressive model with no intervention capacity or outcome telemetry is unlikely to realize quality value.

3. Medication Adherence: Stronger Evidence, Narrower Attribution

Earlier reviews documented AI-enabled prediction, monitoring and adherence-support applications across heterogeneous populations and technologies.[7] The evidence now includes a more directly Medicare-relevant program. Worrall and colleagues evaluated an AI-supported, pharmacist-led medication-adherence service across 10,477 patients using a multicenter retrospective quasi-experimental pre/post design.[19]

After implementation, adherence improved across the three medication categories that remain in the 2027 Medicare Part D Star Ratings: hypertension by 5.9%, cholesterol by 7.9% and diabetes by 6.4%. The investigators also reported that Medicare Star ratings improved.[19,26]

This materially strengthens the measure-level evidence but does not establish that AI alone caused the improvement. The intervention combined AI-supported analytics, case review and pharmacist outreach, and the study was not randomized.

Randomized evidence also shows why scale decisions should remain selective. REINFORCE found a 13.6-percentage-point adjusted improvement in medication adherence with reinforcement-learning-personalized messaging in a small diabetes population.[6] By contrast, Ho and colleagues randomized 9,501 patients across three U.S. health systems and found that generic reminders, behavioral nudges and nudges plus a fixed-message chatbot did not produce sustained 12-month refill-adherence improvement after correction for multiple comparisons; clinical events also did not differ.[21]

> **Executive implication**
> Medication adherence is reasonable for controlled scaling or structured pilots when AI is embedded in a pharmacist or care-management workflow. Require measure-level outcomes, a comparator, sustainability tracking and enough process data to separate the value of targeting from the value of the human intervention.

4. Care Gaps: Actual Completion Changes the Evidence Standard

Large-scale prediction studies have shown that machine learning can improve quality-measure targeting, but earlier estimates of gap closure were sometimes simulated rather than observed.[5] The 2026 Park study moves the evidence to completed preventive care.

Park and colleagues evaluated a real-world machine-learning-guided colorectal-cancer screening program. The model identified higher-risk patients overdue for screening, and a care-gap nursing team conducted personalized outreach. Using a regression-discontinuity design around the model's risk threshold, the program increased colonoscopy uptake by 6.0 percentage points at three months and 6.9 percentage points at six months.[20]

The appropriate causal object is the ML-guided program, not the model alone: the algorithm determined who was targeted and human nurses delivered the intervention. The study did not report contract-level HEDIS or Medicare Advantage Star movement.

> **Executive implication**
> For care-gap vendors and internal models, require actual completion as the success endpoint. Model lift, risk stratification and simulated closure should not be accepted as evidence that a gap was closed. Scaling also requires enough nurse/outreach capacity to act on the model's prioritization.

5. Member Engagement: Personalization Is Not the Product

The engagement evidence argues against a simple 'more personalization equals better outcomes' strategy. REINFORCE demonstrates that adaptive personalization can improve medication-taking behavior in some settings.[6] The much larger Ho trial shows that reminders, behavioral nudges and fixed-message chatbot augmentation do not consistently produce durable adherence gains.[21]

Channel choice is similarly context-dependent. A 2026 retrospective analysis found phone outreach generally produced higher preventive-care compliance than chatbot or multichannel approaches, although chatbot outreach performed better in one diabetes comparison.[8] Other electronically delivered preventive-care outreach has also been evaluated against actual service completion rather than engagement alone.[9]

QII synthesis — intervention orchestration

Personalization is not itself an intervention effect. The operational problem is intervention orchestration: who should receive outreach, through which channel, at what intensity and time, and what should happen when a low-intensity digital intervention does not resolve the underlying barrier.

Executive implication

Do not buy 'personalization' as a feature claim. Buy or build a testable intervention strategy with a completion endpoint, channel-selection logic and an escalation path when digital messaging is insufficient.

6. Population Health: Separate Adoption, Effectiveness and MA Applicability

The population-health evidence is now strong enough to support three different statements—but not to collapse them into one.

Evidence question	What the evidence shows	Boundary
Payer adoption	NAIC's 2025 Health Insurance AI/ML Survey documents insurer-reported production use of AI/ML for Stars-related care gaps, chart abstraction, risk/severity prioritization, proactive quality outreach and identification of gap-closure evidence.[24]	Deployment is real; adoption does not prove effectiveness.
Contemporary effectiveness	Across 48 UCLA Health clinics, AI-identified proactive care management for 3,007 high/highest-risk patients was associated with a 27% level reduction in potentially preventable admissions; preventable ED visits did not show the same significant reduction.[23]	The treatment is AI plus high-touch care management, not algorithm alone.
Medicare Advantage applicability	A payer program for high-risk MA members with CHF used predictive targeting and was associated with lower hospital/ED utilization, faster PCP follow-up and $716 lower total medical cost per month during the first six months after outreach.[22]	MA applicability is demonstrated, but the study uses older predictive analytics and bundled outreach.
Implementation context	Earlier integrated-system predictive-analytics work linked risk-based targeting to standardized care coordination and lower readmission.[10]	Supports the importance of workflow, not a universal AI effect.

Executive implication

Use AI where it changes allocation of scarce clinical or outreach capacity and where downstream utilization or quality outcomes can be measured. Do not use payer adoption statistics as a substitute for effectiveness evidence.

7. Risk Is Not the Same as Likelihood of Benefit

A high-risk member is not necessarily the member most likely to benefit from the intervention an organization can actually deliver. That distinction matters whenever nurse, pharmacist, social-work or outreach capacity is limited.

Causal machine-learning methods are designed to estimate individualized treatment effects rather than only outcome probabilities.[11] The studies in this briefing show the operational relevance of that distinction: Park used risk scores to allocate limited nurse outreach capacity,[20] Raldow used AI-identified risk to trigger high-touch care management,[23] and REINFORCE adapted message framing over time.[6]

These studies do not establish individualized treatment-effect modeling as standard practice in MA. They do support QII's view that intervention-benefit targeting is a potentially higher-value frontier when multiple feasible interventions compete for limited capacity.

Executive implication

If two members have similar predicted risk but only one is likely to respond to the intervention you can deliver, a pure risk score can misallocate scarce capacity. Advanced programs should increasingly test whether intervention-benefit estimation outperforms risk ranking for resource allocation.

8. Digital Quality: AI Can Become Part of the Improvement Loop

Digital quality infrastructure remains the foundation. NCQA describes Digital HEDIS as standardized, computable FHIR/CQL measures with patient-level outputs that can integrate into reporting, analytics, care-gap workflows and population-health programs.[12]

The 2026 Boussina cluster randomized trial demonstrates a more consequential use: AI-enabled medical-record abstraction coupled with near-real-time feedback. Sixty-six emergency physicians caring for 301 SEP-1-eligible patients were randomized to the intervention or conventional feedback. SEP-1 compliance increased from 70.1% to 82.9%, an adjusted absolute improvement of 13.0 percentage points, and LLM-expert agreement was 92%.[25]

The trial did not show improvement in ICU admission or 30-day mortality and involved hospital SEP-1 rather than HEDIS or MA Stars. The evidence therefore supports a process-measure and feedback-loop claim, not a patient-outcome or HEDIS/Star claim.

Executive implication

For MA leaders, the transferable opportunity is faster measurement-to-feedback cycles. Pilot AI-enabled abstraction where it can materially shorten the time between performance and corrective action—but validate the intended HEDIS/Star workflow rather than assuming SEP-1 results will generalize.

9. Documentation AI: Valuable Workflow Tool, Not Yet a Quality Claim

Ambient documentation AI illustrates why operational value and quality value should be separated. A pragmatic randomized trial of 238 outpatient physicians found product-specific effects on time-in-note, with one ambient scribe reducing time by 9.5% while another did not show statistically significant time savings; clinicians also reported occasional clinically significant inaccuracies.[13]

A separate 2026 randomized orthopedic study reported workflow benefits but identified 19 medical errors in AI-scribe notes compared with nine in conventional documentation, with the errors identified before finalization.[14]

Executive implication

Buy documentation AI for documentation or workforce value unless a separate quality pathway is tested. If the business case is HEDIS/Stars, require evidence linking the documentation change to usable quality data, measure capture or clinical action and then to the targeted measure.

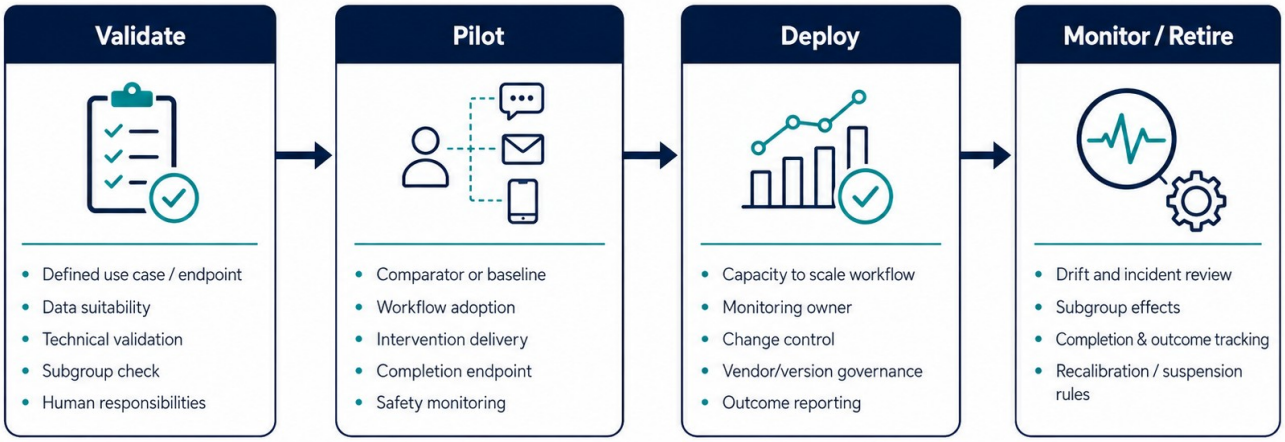
10. Governance Should Operate as Stage Gates

NIST's AI Risk Management Framework organizes risk-management activity around Govern, Map, Measure and Manage; as of August 2026, AI RMF 1.0 remains the published voluntary framework and NIST states that it is being revised.[15] ASTP/ONC's Decision Support Interventions materials provide a healthcare-specific example of transparency and performance expectations within the applicable certified-health-IT scope.[16]

QII recommends translating those principles into stage gates that determine whether a quality-AI use case is ready to advance.

Figure 5. QII Quality-AI Deployment Stage Gates

QII recommended decision gates for progressing a quality-AI use case



QII operating recommendation — not a direct reproduction of NIST or ASTP/ONC.

QII stage gate	Decision point	Minimum evidence / control	Stop rule
VALIDATE	Before pilot	Defined use case and endpoint; source/data suitability; technical validation; subgroup assessment where relevant; human/workflow responsibilities; failure modes.	Do not pilot if the intended action or success endpoint is undefined.
PILOT	Controlled real-world use	Comparator or defensible baseline; workflow adoption; intervention delivery/fidelity; completion endpoint; safety/quality monitoring; override/escalation rules.	Do not scale on model metrics or user enthusiasm alone.
DEPLOY	Broader operating use	Evidence that the workflow can deliver the intended intervention at required capacity; monitoring ownership; change control; vendor/version governance; outcome reporting.	Do not declare HEDIS/Star or ROI success beyond the level actually measured.
MONITOR / RETIRE	Postdeployment	Technical drift, workflow drift, safety, subgroup effects, intervention completion, quality/economic outcomes, incident response, recalibration/suspension thresholds.	Retire or redesign when the human-AI program no longer produces acceptable value or risk.

Executive implication

Govern the human-AI program, not only the model. The model, data, workflow, intervention capacity, human decision rights, escalation path and downstream outcome monitoring are all part of the system producing—or failing to produce—quality value.

11. Match the Evidence to the Claim

QII's evidence standard is simple: the evidence should reach the same level as the claim. This is consistent with the separation between prediction-model reporting, early live AI evaluation and AI intervention trials reflected in TRIPOD+AI, DECIDE-AI and CONSORT-AI.[2–4]

Claim	Not enough by itself	Evidence expected
'The model identifies quality risk.'	Anecdotes or user satisfaction	Technical validation appropriate to the intended use, including calibration/discrimination and relevant subgroup performance.
'The AI improves workflow.'	Model accuracy alone	Observed time, burden, usability, fidelity, error/override and workflow-performance evidence.
'The AI closes care gaps.'	Risk ranking or simulated closure	Observed intervention completion or gap closure with a defensible comparison/attribution approach.
'The AI improves quality.'	Clicks, alerts, adoption or documentation volume	The relevant process, quality or clinical endpoint at the level claimed.
'The AI improves equity.'	Aggregate model performance	Relevant subgroup performance plus intervention access/completion and downstream outcome analysis.
'The AI creates ROI.'	Projected savings	Observed benefits, relevant implementation/operating costs, attribution and time horizon.

11.1 Intervention Fidelity Chain™

Adoption is an implementation outcome, but it is not proof that the intended intervention occurred. QII uses the Intervention Fidelity Chain™ to separate generated → delivered → viewed → accepted/selected → initiated → completed. The exact telemetry required should be proportionate to the use case.

11.2 Equity Evidence Crosswalk™

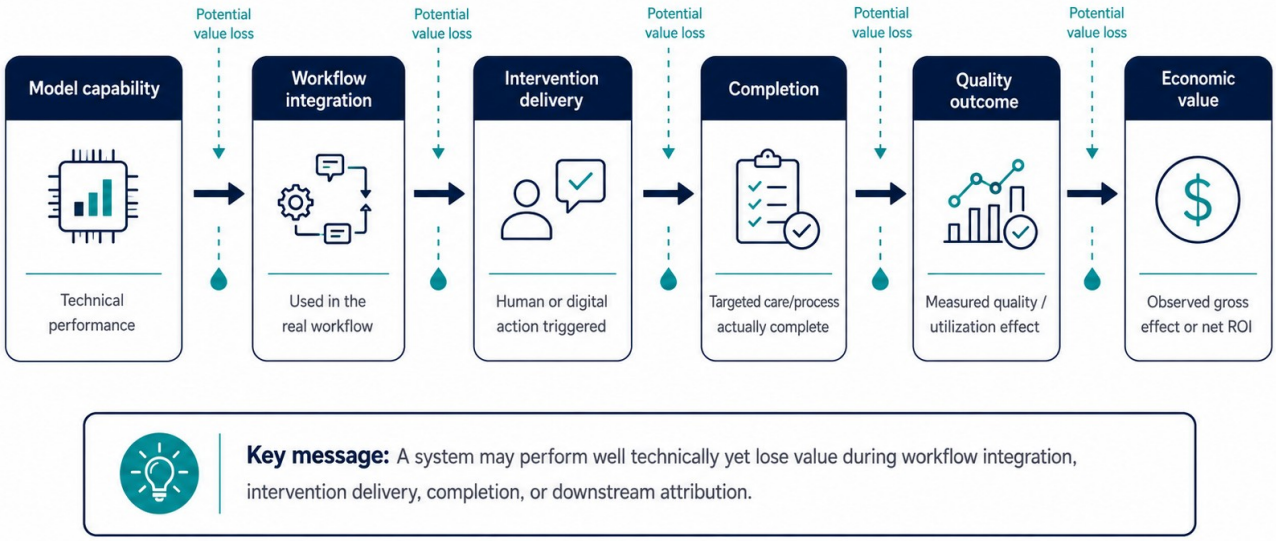
Equity should also be followed downstream. QII separates representation in the data, subgroup model/system performance, intervention access and completion, and downstream outcomes. A favorable fairness metric at one stage should not be treated as evidence of equitable real-world outcomes. Obermeyer and colleagues demonstrated how a proxy target can produce racially biased allocation despite apparent operational usefulness. [18]

11.3 AI Value Realization Gap™

The updated evidence narrows—but does not eliminate—the AI Value Realization Gap™. Actual intervention completion, CMS process-measure improvement, utilization change and gross medical-cost effects are now documented. The largest remaining gaps are AI-specific attribution to contract-level HEDIS/Star performance and attributable net economic return.

Figure 4: AI Value Realization Gap™

Illustrative flow showing where value can be lost



12. Economic Value: Report Three Different States

The economic evidence is no longer accurately summarized as simply weak. Ukert and colleagues found $716 lower total medical cost per month during the first six months after predictive-analytics-guided outreach in high-risk MA members, alongside lower hospitalization and ED utilization.[22]

That is a realized gross economic effect, not a complete AI ROI calculation. A 2026 systematic review of 117 medical-AI economic evaluations found generally suboptimal reporting; 63% of studies evaluated relatively early-readiness systems, only 28% included implementation costs and 57% reported operational costs.[17] QII recommends reporting economic value in three distinct states:

Economic state	Meaning	Executive use
Projected value	Scenario or business-case estimate before the outcome occurs.	Useful for prioritization; should not be reported as realized savings.
Observed gross economic effect	Measured medical or operational change associated with the program.	Evidence of economic effect; may still combine AI, human intervention and workflow effects.
Net AI ROI	Attributable benefits minus the technology, integration, implementation, intervention labor, monitoring, governance and ongoing operating costs required to produce them.	Highest evidentiary bar; currently underdeveloped for mature MA quality AI.

Executive implication

Finance should receive projected value, observed gross effect and net ROI as three different states. Do not collapse them into one 'AI savings' number.

13. What Medicare Advantage Executives Should Do Now

QII recommendation	Decision consequence
1. Start with the intervention system	Define the measure/problem, target population, model role, human action, workflow bottleneck, capacity constraint and success endpoint before selecting technology.

QII recommendation	Decision consequence
2. Require actual completion for care-gap claims	Do not accept AUROC, lift or simulated closure as proof that care changed.
3. Test personalization rather than assuming it	Require a durable completion/adherence endpoint and an escalation path beyond low-intensity digital messaging.
4. Separate adoption, effectiveness and MA applicability	Use regulator deployment data to understand adoption; use comparative outcome evidence for effectiveness; use MA-specific evidence for MA claims.
5. Treat measurement speed as an intervention lever	Where the measure supports it, test whether AI-enabled abstraction shortens the feedback cycle enough to change performance.
6. Govern through stage gates	Validate → Pilot → Deploy → Monitor/Retire, with evidence requirements at each transition.
7. Distinguish workflow ROI from quality ROI	Documentation and productivity benefits are legitimate; do not relabel them as HEDIS/Star value without a tested pathway.
8. Distinguish gross economic effect from net AI ROI	Include the full cost stack and attribution method before reporting return on investment.

What this means for Stars leaders

Treat patient-level adherence improvement, screening completion and hospital process-measure improvement as important evidence—but not as automatic contract-level Star lift. For a Stars business case, require a defined pathway from the AI-enabled program to the relevant measure population, numerator/denominator effect, cut-point or performance movement, and contract-level attribution.

Conclusion

The evidence base for AI in healthcare quality has moved beyond prediction and prototypes. The reviewed literature now includes actual screening completion, medication-adherence improvement, payer production deployment, selected utilization reduction, Medicare Advantage medical-cost effects and randomized improvement in a CMS process measure.

The most defensible interpretation is not that AI independently produces these outcomes. It is that AI can create measurable value when embedded in a functioning intervention system with usable data, defined human actions, operational capacity, timely feedback and outcome measurement.

QII's decision standard for Medicare Advantage quality is therefore straightforward: do not ask only whether the model is accurate. Ask whether the organization can act on the output, complete the intervention, measure the quality effect at the level claimed, and demonstrate value after the full cost of making the human-AI system work.

Methods and Limitations

This manuscript was developed as a structured QII evidence review under QII Research Standards v1.0. It is not a systematic review or meta-analysis. The process used a predefined protocol, Formal Source Set v1.2, RB-001 Evidence Table v0.4.1, source screening, targeted contradictory-evidence searches, independent claim-level verification and current-source/citation review for material claims.

The Formal Source Set contains 54 screened/included sources spanning CMS, NCQA, NIST, ASTP/ONC, NAIC and peer-reviewed literature. Twenty-four controlled claims (MAQ-C001–MAQ-C024) are maintained in Evidence Table v0.4.1. Full-text or primary-source verification was prioritized for material claims and numerical findings. QII did not conduct a comprehensive database search sufficient to claim literature exhaustiveness, and no single formal study-level risk-of-bias instrument was applied across the entire source set.

Important residual limitations remain. Worrall closes the Medicare-relevant adherence/Star measure-level gap but does not isolate AI's independent contribution. Park demonstrates actual ML-guided preventive-service completion but not contract-level HEDIS/Star improvement. Ukert establishes gross medical-cost effects in an MA

predictive-analytics program but not net AI ROI. Boussina demonstrates a randomized CMS process-measure effect but not HEDIS-specific performance or patient-centered benefit. NAIC documents insurer production deployment, but adoption is not effective.

Evidence and current regulatory/program sources were refreshed through August 2026. The final claim-to-citation reconciliation against Evidence Table v0.4.1 identified no substantive evidence-boundary failures and was completed before editorial freeze. Final release QA was completed before public release.

References

1. Centers for Medicare & Medicaid Services. Contract Year 2027 Medicare Advantage and Part D Final Rule. April 2, 2026.
2. Collins GS, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378.
3. Vasey B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. 2022;377:e070904. doi:10.1136/bmj-2022-070904.
4. Liu X, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nature Medicine. 2020;26(9):1364-1374. doi:10.1038/s41591-020-1034-x.
5. Patel SY, Barnett ML, Basu S. Predicting quality measure completion among 14 million low-income patients enrolled in Medicaid. npj Digital Medicine. 2025.
6. Lauffenburger JC, et al. The impact of using reinforcement learning to personalize communication on medication adherence: findings from the REINFORCE trial. npj Digital Medicine. 2024;7:39. doi:10.1038/s41746-024-01028-5.
7. Babel A, Taneja R, Mondello Malvestiti F, Monaco A, Donde S. Artificial Intelligence Solutions to Increase Medication Adherence in Patients With Non-communicable Diseases. Frontiers in Digital Health. 2021;3:669869. doi:10.3389/fdgth.2021.669869.
8. Chacko J, et al. Chatbot Outreach in Value-Based Preventive Care: Retrospective Analysis. JMIR Medical Informatics. 2026;14:e81370.
9. Lesser LI, Datta E, Behal R. Electronically-Delivered Push Notifications Improve Patient Adherence to Preventive Care. Journal of the American Board of Family Medicine. 2025;38(2):239-246. doi:10.3122/jabfm.2024.240088R1.
10. Marafino BJ, et al. Evaluation of an intervention targeted with predictive analytics to prevent readmissions in an integrated health system: observational study. BMJ. 2021;374:n1747.
11. Feuerriegel S, et al. Causal machine learning for predicting treatment outcomes. Nature Medicine. 2024;30(4):958-968. doi:10.1038/s41591-024-02902-1.
12. National Committee for Quality Assurance. Digital HEDIS. Current resource accessed August 2026.
13. Lukac PJ, et al. Ambient AI Scribes in Clinical Practice: A Randomized Trial. NEJM AI. 2025. doi:10.1056/AIoa2501000.
14. Etagiuri N, et al. Evaluating the Impact of Artificial Intelligence Scribes on Clinical Workflow and Documentation Quality: A Randomized Controlled Trial. Journal of Arthroplasty. 2026. doi:10.1016/j.arth.2026.03.069.
15. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0) and AI RMF Playbook: Manage. Current web resources accessed August 2026.
16. Assistant Secretary for Technology Policy/Office of the National Coordinator for Health Information Technology. Decision Support Interventions: §170.315(b)(11) certification criterion and test method. Current resource accessed August 2026.
17. Godoy Junior CA, et al. Technological Maturity and Cost-Effectiveness of Medical Artificial Intelligence: A Systematic Review of Health Economic Evaluations. Value in Health. 2026;29(5):896-904. doi:10.1016/j.jval.2026.01.014.
18. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-453. doi:10.1126/science.aax2342.
19. Worrall C, Shirley D, Bullard J, Dao A, Morrisette T. Impact of a clinical pharmacist-led, artificial intelligence-supported medication adherence program on medication adherence performance, chronic disease control measures, and cost savings. Journal of the American Pharmacists Association. 2025;65(1):102271. doi:10.1016/j.japh.2024.102271.
20. Park M, Chan CW, Boell K, Mitchell EG, Tariq AA, Vawdrey DK. Cancer Screening Outreach Guided by Machine Learning: The Benefits of Proactive Care. Manufacturing & Service Operations Management. 2026;28(4):1011-1030. doi:10.1287/msom.2024.1353.
21. Ho PM, Glorioso TJ, Allen LA, et al. Personalized Patient Data and Behavioral Nudges to Improve Adherence to Chronic Cardiovascular Medications: A Randomized Pragmatic Trial. JAMA. 2025;333(1):49-59. doi:10.1001/jama.2024.21739.
22. Ukert B, David G, Smith-McLallen A, Chawla R. Do payor-based outreach programs reduce medical cost and utilization? Health Economics. 2020;29(6):671-682. doi:10.1002/hec.4010.
23. Raldow AC, et al. Proactive Care Management of AI-Identified At-Risk Patients Decreases Preventable Admissions. American Journal of Managed Care. 2024;30(11):548-554. doi:10.37765/ajmc.2024.89625.
24. National Association of Insurance Commissioners. Health Insurance Artificial Intelligence/Machine Learning Survey Results. 2025.
25. Boussina A, Allison C, Quintero K, et al. Medical Record Abstraction for Quality Improvement in Sepsis Care Using Artificial Intelligence: A Cluster Randomized Trial. JAMA Network Open. 2026;9(6):e2611885. doi:10.1001/jamanetworkopen.2026.11885.
26. Centers for Medicare & Medicaid Services. 2027 Star Ratings Measures and Weights. 2026.

QUALITY INTELLIGENCE INSTITUTE

DEFINING THE FUTURE OF HEALTHCARE AI QUALITY.

QII Research Briefing #1

The State of AI in Medicare Advantage Quality

PUB-RB-001 | Final Release v1.1 — Brand Controlled

qualityintelligenceinstitute.org