PROCESS · MEASURING VALUE

Measure AI Value Before and After a Pilot

An OmniTek Consulting implementation guide for VPs and Directors in pharma, biotech, and medical device companies.

READING TIME
About 5 minutes

WHO IT'S FOR
VPs and Directors

GOES WITH
The AI Maturity Quick Check

> AI projects that can't show results are the first to lose funding. Measuring one or two metrics before a pilot starts, and the same metrics after, takes a few hours and gives you a number your executives will trust.

SCOPE This guide covers AI used in how your company works, such as drafting, analysis, and operations. It doesn't cover AI that is part of a product you sell, such as AI-enabled device software or software as a medical device. That has its own regulatory path; check it with your Regulatory team.

Why measure before you start

- **You can't recreate a baseline later.** Once the pilot is running, nobody remembers how long the task used to take.
- **Stories don't survive budget season.** "It saves a lot of time" loses to a specific figure, such as "report preparation dropped from 12 hours to 5" (illustrative numbers).
- **Quality matters as much as speed.** In regulated work, a faster process with more rework or errors is not a win. Measure both.

Metrics by department

Pick one speed metric and one quality metric per pilot.

Department	Speed metric	Quality metric
Quality	Days to close a deviation; hours to prepare the monthly quality report	Investigations reopened; report corrections after review

Department	Speed metric	Quality metric
Clinical Operations	Days from monitoring visit to letter; time to database lock	Queries raised on AI-assisted documents; protocol deviation trends
Regulatory Affairs	Hours to draft a response or submission section	Review comments per section; rounds of revision
Pharmacovigilance	Case processing time	Case quality review findings; late cases
Manufacturing	Batch record review time	Right-first-time batch records; review findings
Medical Affairs	Time to answer a medical information request	Responses needing correction after review

Also track one simple adoption measure, such as how many of the team used the tool for this task in the last month.

A pilot measurement plan

One page per pilot, filled in before it starts.

TEMPLATE · COPY AND ADAPT

Pilot Measurement Plan: [AI use]

Owner: [ ] | Pilot dates: [start] to [end] | GxP class: [ ]

What we're measuring

Speed metric: [e.g. hours to prepare the monthly quality report] Quality metric: [e.g. corrections requested in review] Adoption metric: [e.g. share of eligible reports drafted with the tool]

How we'll measure

Source: [time log, system timestamps, review records] Baseline period: [last 3 cycles or 4 weeks before the pilot] Who records it: [ ]

Baseline (before)

Speed: [ ] Quality: [ ] Volume in the period: [ ]

Success looks like

[e.g. at least 30% faster with no increase in corrections (illustrative target)]

A results summary

At the end of the pilot, report on one page in the same order.

	Before	After	Change
Speed metric	[]	[]	[]
Quality metric	[]	[]	[]

	Before	After	Change
Adoption	n/a	[ ]	[ ]

Then add three short lines: what worked, what didn't, and the recommendation (scale, adjust, or stop). Convert time saved into hours per year for the whole team, which is the number leadership will remember.

Your first 30 days

1. **Week 1: Pick the pilot and the metrics.** Choose one speed and one quality metric from the table above, or your own.
2. **Week 1: Find the data.** Check whether system timestamps or review records already capture the metric. If not, set up a simple time log. Pulling timestamps for a pilot measurement is operational reporting: low risk as long as no GxP decision relies on the extract. Handle the extract as your data integrity SOP requires, and never use it to show a GxP limit (such as a hold time) was met.
3. **Week 2: Record the baseline.** Measure the last few cycles, or the next few before the tool is used.
4. **Weeks 3 and 4: Run the pilot and keep measuring.** Use the same method and the same people as the baseline. If the pilot touches GxP work such as batch record review or case processing, classify the use first; GxP direct uses need validation before a pilot runs on live records, so pilot those in a test environment kept separate from validated systems, using copies handled as your data integrity SOP requires.
5. **End of the pilot: Report.** Fill in the results summary and share it with your sponsor, whether the result is good or not.

By the end of your first pilot you have a measured result to report. Doing this for every AI use, and reporting regularly, covers the value questions in the Quick Check at the Operationalizing level and above.

Common mistakes

- **No baseline.** The most common gap, and the one you can't fix afterward.
- **Only measuring speed.** In regulated work, faster with more errors is a step backward. Always pair speed with quality.
- **Changing the method mid-pilot.** Measure the same way before and after, or the comparison means nothing.
- **Counting the enthusiasts only.** Early users are often the fastest. Measure across the people who will use it day to day.
- **Hiding a weak result.** A pilot that didn't work, reported honestly, builds more credibility than a vague success.

How OmniTek can help

The full AI Maturity Report Card measures how well your department tracks AI value, and its 30/60/90-day plan includes the metrics to use for your top gaps.

To talk through how to measure your first pilot, or your Quick Check results, book time with OmniTek.

Book time with OmniTek

This guide is general information for leaders in regulated life sciences functions, not legal or regulatory advice. Confirm requirements with your Quality, Regulatory, and Legal teams. © 2026 OmniTek Consulting.