Implementation Guide

See, Diagnose, Redesign, Test, Measure, Scale, Sustain. How to run it in your setting, and how it usually fails.

> **See › Diagnose › Redesign › Test › Measure › Scale › Sustain.**
> The first six are the sequence used in the session. **Sustain is the seventh, and it is the one most programs skip.** This sequence does not replace your improvement method. Use it with PDSA, the Model for Improvement, Lean, Lean Six Sigma or whatever your organization already runs. Its only job is to make workforce design visible inside QI work.

01 SEE

Start with observation, not diagnosis.

Portal messages taking longer to close; more after-hours EHR work; resident continuity declining; turnover rising; patient response time worsening; tasks escalated inconsistently.

> Avoid premature diagnoses: “we are understaffed”, “people are not resilient”, “we need AI”, “the MAs need to do more”.

02 DIAGNOSE

Use the five domains to decide where to look.

Capacity, Control, Workflow, Team, Leadership. Ask which appears most relevant and what evidence would confirm or challenge that.

> Score the domains separately and never total them. Role-to-role disagreement is a useful signal for further inquiry.

03 REDESIGN

Select one modifiable condition.

Standardize message routing; clarify escalation criteria; protect review time; remove duplicate review; rebalance a panel or schedule template; define role ownership; retire a low-value requirement.

> Prefer one primary modifiable condition in the first test. When multiple changes are operationally inseparable, document them as a bundled intervention and be explicit that the test cannot isolate which component produced the result.

04 TEST

Keep the first test small.

One pod. One clinician. One workflow. Two weeks. One message category. One huddle process.

> Small is not timid. Small is what makes the result interpretable and the failure survivable.

05 MEASURE

Four categories, and the fourth is mandatory.

Workforce measure; operational or process measure; care or QI measure; balancing measure.

> Reducing physician burden by moving it to nursing is burden transfer, not improvement. The balancing measure is what catches it.

06 SCALE

Spread it only when the evidence justifies spreading it.

Scale after the test produces enough evidence to warrant expansion. If the result is mixed, adapt and retest. If it created burden transfer or a safety risk, stop or redesign.

> A mixed result is information, not permission. Scaling on a mixed result is how a good idea acquires a bad reputation.

07 SUSTAIN

Hold the gain. This is the step most programs skip.

Name what becomes standard work and who owns it. Keep one measure on a dashboard with a named reviewer and a cadence. Define the signal that says it is slipping, and what happens then.

> Spreading and holding are different jobs. An improvement with no owner and no measure is a pilot that will quietly revert. Decide what stopped or simplified to make room for it.

Applying it in different settings

The five domains do not change between settings. What changes is which ones you can actually move, and who needs to be in the room.

Residency and academic practice

Watch for role ambiguity between residents, faculty, nurses, medical assistants and supervising clinicians. **Not every handoff is waste:** some supervision and escalation are clinically and educationally required, and a redesign that removes them is a patient-safety and accreditation problem, not an efficiency gain. Align any test to the block schedule, because a 30-day test spanning a block change is measuring two different systems.

FQHC

Demand is often not discretionary and payer mix limits the capacity response. When headcount is constrained, Capacity may be less immediately modifiable, and Workflow and Team can sometimes offer more reachable first tests: routing accuracy, first-touch resolution, standing orders and protocol pathways within approved scope. Balancing measures matter more here than anywhere, because support staff have the least slack and burden transferred to them shows up fastest.

Small independent practice

Control is often the strongest domain because the decision-maker may be in the room and a change can happen on Monday. What is missing is analytics, spare capacity and a second pod to compare against. So shrink everything: count by hand and run it on one clinician for two weeks. For an initial local test a small hand-count may be enough to reveal an obvious process signal. Match the amount of data to the decision you are trying to make, be explicit that it is a before-and-after rather than a controlled comparison, and do not imply statistical inference from a small convenience sample.

Large integrated system

The data usually exists, but the person who controls the condition may be three levels away in a different reporting line, and standard workflows may be centrally governed. Identify early which of the five domains sit inside local authority. Testing a centrally governed workflow without the necessary sponsorship is a common implementation failure mode. Win locally, then use that result to earn the conversation about the governed one.

When capacity truly is insufficient

Sometimes it is, and saying so plainly protects your credibility. Redesign usually strengthens the staffing case rather than replacing it: demand-versus-capacity data, touches per message and after-hours work are much harder to decline than how people feel. But if demand exceeds capacity by a wide margin, workflow improvement will not close the gap. Quantify it and move the decision to where it belongs.

Ten questions to take back

1. Where is demand growing faster than our operating model?
2. Where are people accountable for outcomes they have little ability to influence?
3. Where does work routinely wait, loop, duplicate or follow people home?
4. Where does role ambiguity create unnecessary escalation or rework?
5. What recurring problem have we normalized rather than redesigned?
6. Which performance measures look stable only because people are compensating?
7. What work should stop, simplify or move before we add something new?
8. What is one small change we could test in the next 30 days?
9. How will we know whether we improved both workforce experience and system performance?
10. What balancing measure will tell us if we merely transferred the burden?

Common ways workforce-design improvement efforts lose traction

Named plainly, because recognising them early is cheaper than discovering them at month six.

- **Intervening before diagnosing.** Every domain has an obvious-sounding fix, and most obvious fixes add work to the people already carrying the most.
- **Transferring the burden.** The physician measure improves, RN workload explodes, and the dashboard calls it a success.
- **Testing a governed workflow without sponsorship.** A common way this work dies in a large system.
- **Measuring only burnout.** That measures people's response to the system without measuring the system.
- **Scaling on a mixed result.** Adapt and retest instead. A mixed result is information, not permission.
- **Skipping Sustain.** No owner, no measure, no cadence, and the gain is gone by the next block change.

Disclosure. Dr. Bains is Founder and CEO of Energized Vision. He has no relevant financial relationships with ineligible companies to disclose. The STFM session this accompanies is educational and is not a commercial presentation.

Suggested citation. Bains A. From Burnout to System Resilience: Workforce Design Strategies That Improve Quality and Safety. Workforce Design Toolkit. Energized Vision; 2026. energizedvision.org/stfm

Version 1.0 · August 2026