

Three ways to change what a physician orders. Only two of them work.

And what happens when you try the same thing on imaging.

Most imaging software implements the one that didn't. Including, until recently, ours.

We also said, until this month, that nobody had studied this on imaging. They had. We were wrong and this page is the correction.

Trial figures — Source Register / Vol III §5.1 · mechanisms.json v1.0.0 · [measured]

1. What we know about imaging

The imaging evidence is the Duke Primary Care peer-comparison dashboard, reported on two denominators — volume and cost — not two independent findings. Baseline top-decile:bottom-decile variation 4.2×; year-one median imaging rate −17.3% (within-network pre-post). Median estimated radiology costs −19.5%; >\$3M saved in year one (within-network pre-post). Design: observational, no control group, no risk adjustment, effect not separable from secular trend. These are within-network pre-post changes; an unknown share is secular trend.

OBSERVATIONAL · NO CONTROL GROUP

At Duke, the top decile of primary-care providers ordered 4.2× more imaging than the bottom decile — before anyone intervened. [measured]

Study	Effect	Design (qualifier)	GRADE
JGIM volume result	year-one median imaging rate −17.3% (within-network pre-post)	OBSERVATIONAL · NO CONTROL GROUP	GRADE LOW
JACR cost result	Median estimated radiology costs −19.5%; >\$3M saved in year one (within-network pre-post)	OBSERVATIONAL · NO CONTROL GROUP	GRADE LOW

How the studies compare

Study	Domain	Design	Control	GRADE
halpern-2021	imaging	observational pre-post	No	low
clark-randall-2021	imaging	observational pre-post	No	low
meeker-2016	analogous-domain (not imaging)	cluster-randomised clinical trial	Yes	moderate
linder-2017	analogous-domain (not imaging)	prespecified secondary analysis of a cluster-RCT	Yes	low

2. The four-arm trial — raw drop vs difference-in-differences

Meeker et al., *JAMA* 2016 — cluster-randomised across 47 primary-care practices, 248 clinicians, 18 months. Do not quote the raw within-arm drop: the control arm fell 11.0 percentage points on secular trend alone. The honest attributable effect is the difference-in-differences (about 5–7 points for the arms that worked). Suggested alternatives had the large raw drop everyone quotes — and was not significant vs control. That is the arm most imaging software ships.

Arm	Raw change	DiD vs control	ARKA
Control Education only	24.1% → 13.1% (–11.0 pts)	—	—
Suggested alternatives EHR order set offering non-antibiotic options at the point of prescribing	22.1% → 6.1% (–16.0 pts)	–5.0 pts (95% CI –7.8 to 0.1), p = 0.66 — not significant	As a pre-built default, never as a suggestion
Accountable justification Free-text justification typed into the patient's record when prescribing against guideline	23.2% → 5.2% (–18.0 pts)	–7.0 pts (95% CI –9.1 to –2.9), p = 0.001 — significant	Module 3 — with a verified record write
Peer comparison Monthly email comparing the clinician's own rate to the lowest-rate top performers — no EHR component	19.9% → 3.7% (–16.2 pts)	–5.2 pts (95% CI –6.9 to –1.6), p = 0.001 — significant	Module 2 — monthly, outside the EHR

3. Twelve-month persistence after removal

Peer comparison also decayed after the interventions stopped — but it was the sole arm still significantly below control twelve months later. Accountable justification produced the largest effect during the intervention and was statistically indistinguishable from control a year after it was removed.

Arm	End of intervention	12 months post	DiD vs control
Control	14.2%	11.8% (–2.4)	—
Suggested alternatives	7.4%	8.8% (+1.4)	+3.8, p = 0.55
Accountable justification	6.1%	10.2% (+4.1)	+6.5, p = 0.001
Peer comparison	4.8%	6.3% (+1.5)	+3.9, p = 0.005

4. Wider evidence (Vol III §5.4)

- **Peer comparison also moves statin prescribing in primary care — the mechanism is not antibiotic-specific.**

Imaging ordering is a fair extension of the same behavioural channel; it has not yet been trialled inside a risk-bearing population with persistence measured after withdrawal.

Persell SD, Doctor JN, Friedberg MW, et al. Behavioral Interventions to Increase Statin Prescribing: A Cluster Randomized Clinical Trial. *JAMA Cardiol.* 2016. [measured]

- **The settled nudge taxonomy separates defaults, social norms, and justification — they are not interchangeable.**

Suggested alternatives (defaults / order sets) are the arm that failed vs control. Building only that layer is implementing the null result.

Thaler RH, Sunstein CR. Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale University Press; 2008. [modeled]

- **Clinicians' self-estimates of their own CT utilisation are systematically inaccurate relative to measured rates.**

Peer comparison works partly because it replaces a biased self-view with an external, risk-adjusted number.

Clinicians systematically mis-estimate their own advanced-imaging utilisation relative to peers — a premise of peer-comparison design (Vol III §5.4; assembled for imaging). [modeled]

- **Status-quo bias makes "do nothing different" the default clinical decision even when evidence favours change.**

A monthly packet that names a specific alternative pathway addresses the bias; a silent CDS alert does not.

Samuelson W, Zeckhauser R. Status quo bias in decision making. J Risk Uncertain. 1988;1:7–59. [measured]

- **Small, frequent incentives ("peanut effect" / mental accounting) change behaviour differently from large lump-sum bonuses.**

Module 5 incentive design must respect mental accounting — do not assume a year-end shared-savings cheque substitutes for monthly peer feedback.

Thaler RH. Mental accounting and consumer choice. Mark Sci. 1985;4(3):199–214. ("peanut effect" / mental accounting in incentive design). [modeled]

- **Defaults save lives in organ donation — and they can also lock in low-value imaging pathways when the default order set is wrong.**

Under-used defaults are a design opportunity; over-trusted defaults are how suggested-alternatives software ships the arm that did not work.

Johnson EJ, Goldstein D. Medicine. Do defaults save lives? Science. 2003;302(5649):1338–1339. [measured]

- **Audit and feedback produces a small-to-moderate, highly variable effect (Cochrane).**

Quote this before a buyer finds it. Peer comparison is the arm that worked and, though it also decayed after removal, remained significantly below control at twelve months — not a guarantee of a large effect in any given organisation.

Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice. Cochrane Database Syst Rev. CD000259. [measured]

- **Brehaut's 15 suggestions: recommend a specific action the recipient controls; align with their own goals; deliver repeatedly; keep key messages short.**

Encoded in the peer-packet schema (two gap cohorts, one named alternative, monthly cadence, outside the EHR).

Brehaut JC, Colquhoun HL, Eva KW, et al. Practice Feedback Interventions: 15 Suggestions for Optimizing Effectiveness. Ann Intern Med. 2016;164(6):435–441. [measured]

5. Settled taxonomy — ARKA's position on each

From docs/arka-aj/mechanisms.json (checksum sha256:41b9c268b...).

Mechanism	Evidence	ARKA status	What we will not do
Accountable justification	cluster RCT · DiD -7 pp · fully decayed by 12 months	Implemented	use the text for performance measurement, credentialing or compensation; store the text
Peer comparison	cluster RCT · DiD -5.2 pp · also decayed — still significantly below control 12 months later	Implemented	rank clinicians; show a bottom-N list with names; attach personal dollar figures to a clinician-facing packet
Suggested alternatives	cluster RCT · DiD -5 pp · Decayed; DiD 3.8%, $p = .55$ — indistinguishable from control	Deliberately not implemented as such	ship a thin draft labelled as a default; claim an effect for a suggestion that still needs clinician input
Defaults and pre-orders	health-system EHR default redesign (quasi-experiment / ITS) · generic prescribing remained elevated ~2.5 years after redesign (Patel follow-up)	Implemented	emit a default that still leaves required fields blank; silently degrade a failed default into a suggestion
Goal setting	programme synthesis (nudge-unit taxonomy)	Planned	pretend a planned mechanism is live; substitute education or alerts for a negotiated goal
Feedback	Cochrane systematic review (audit and feedback)	Implemented	build a justification-analytics dashboard; show clinician-facing ranked lists
Information transparency	nudge-unit taxonomy / status-quo bias literature	Not implemented	claim cost or radiation transparency as a shipped mechanism
Active choice	nudge-unit taxonomy	Deliberately not implemented as such	replace free-text justification with a forced binary or taxonomy choice; gate the order on selecting an option
Alerts and reminders	alert-fatigue / override literature (programme bound)	Deliberately limited	fire above the published ceilings; treat an alert without a record write as accountable justification
Environmental cueing	cluster RCT (public-commitment poster)	Not implemented	omit this mechanism from the ledger to look more complete than we are
Education	cluster RCT control arm · continued secular drift	Not implemented	sell education as the intervention
Hard stops	contractual product commitment — Meeker 2016 used soft behavioural arms, not forced functions	Will never implement	block, delay, or condition an order signature; require justification before the order can proceed

6. ARKA's own correction

The intervention we demonstrated to a former CMS Chief Medical Officer in August 2026 was suggested alternatives. It is the arm of the trial that did not work. We rebuilt. Accountable justification writes into the patient's record and we verify the write; peer comparison arrives monthly outside the EHR; a default

is executable with zero additional clinician input or we label it a suggestion. We publish the fire-rate ceilings and the measured rate. We do not sell the control arm.

7. What a default is — before / after JSON

A default is an order object that requires zero additional clinician input to be signed. If accepting it leaves any required field for the clinician to complete, it is a suggestion, and we label it as one.

Required fields: status === 'draft', intent === 'order', code.coding[] with CPT or LOINC code value (display-only insufficient), subject.reference, requester.reference, encounter.reference, priority ('routine' | 'urgent' | 'asap' | 'stat'), occurrenceDateTime | occurrencePeriod | occurrenceTiming, reasonCode[] non-empty; every entry carries ICD-10-CM, bodySite[] when the study code requires laterality, supportingInfo[] with at least one reference.

A suggestion (before)

```
{
  "resourceType": "ServiceRequest",
  "status": "draft",
  "intent": "order",
  "subject": {
    "reference": "Patient/patient-1"
  },
  "code": {
    "text": "X-ray - lumbar spine",
    "coding": [
      {
        "display": "X-ray",
        "code": "x-ray"
      }
    ]
  }
}
```

A default (after)

```
{
  "resourceType": "ServiceRequest",
  "id": "default-order-1",
  "status": "draft",
  "intent": "order",
  "priority": "routine",
  "subject": {
    "reference": "Patient/p-1"
  },
  "requester": {
    "reference": "Practitioner/prac-1"
  },
  "encounter": {
    "reference": "Encounter/enc-1"
  },
  "occurrenceDateTime": "2026-08-09T15:00:00.000Z",
  "code": {
    "text": "X-ray chest",
    "coding": [
      {
        "system": "http://www.ama-assn.org/go/cpt",
        "code": "71046",
        "display": "Radiologic examination, chest; 2
views"
      }
    ]
  },
  "reasonCode": [
    {
      "coding": [
        {
          "system": "http://hl7.org/fhir/sid/icd-10-cm",
          "code": "R05.9",
          "display": "Cough, unspecified"
        }
      ]
    }
  ],
  "supportingInfo": [
    {
      "reference":
"https://arkahealth.com/evidence/chest-xr"
    }
  ]
}
```

Citations

1. Meeker D, Linder JA, Fox CR, et al. Effect of Behavioral Interventions on Inappropriate Antibiotic Prescribing Among Primary Care Practices: A Randomized Clinical Trial. *JAMA*. 2016;315(6):562–570. PMID 26864410.
2. Linder JA, Meeker D, Fox CR, et al. Effects of Behavioral Interventions on Inappropriate Antibiotic Prescribing in Primary Care 12 Months After Stopping Interventions. *JAMA*. 2017;318(14):1391–1392. PMID 29049577.
3. Halpern DJ, Clark-Randall A, Woodall J, Anderson J, Shah K. Reducing Imaging Utilization in Primary Care Through Implementation of a Peer Comparison Dashboard. *J Gen Intern Med*. 2021;36(1):108–113. PMID 32885372.
4. Clark-Randall A, Halpern DJ, Taylor J, Roth CJ, Gupta RT, Woodall J, Shah KP. Implementation of a Radiology Utilization Dashboard Yields Significant Cost Savings in a Large Primary Care Network. *J Am Coll Radiol*. 2021;18(7):947–950. PMID 33745853.
5. Brehaut JC, Colquhoun HL, Eva KW, et al. Practice Feedback Interventions: 15 Suggestions for Optimizing Effectiveness. *Ann Intern Med*. 2016;164(6):435–441. PMID 26903136.
6. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice. *Cochrane Database Syst Rev*. CD000259.
7. Persell SD, Doctor JN, Friedberg MW, et al. Behavioral Interventions to Increase Statin Prescribing: A Cluster Randomized Clinical Trial. *JAMA Cardiol*. 2016. PMID 27438316.
8. Thaler RH, Sunstein CR. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press; 2008.
9. Samuelson W, Zeckhauser R. Status quo bias in decision making. *J Risk Uncertain*. 1988;1:7–59.
10. Johnson EJ, Goldstein D. Medicine. Do defaults save lives? *Science*. 2003;302(5649):1338–1339. PMID 14631022.
11. Thaler RH. Mental accounting and consumer choice. *Mark Sci*. 1985;4(3):199–214. ("peanut effect" / mental accounting in incentive design).
12. Clinicians systematically mis-estimate their own advanced-imaging utilisation relative to peers — a premise of peer-comparison design (Vol III §5.4; assembled for imaging).
13. Patel MS, Day SC, Halpern SD, et al. Generic Medication Prescription Rates After Health System-Wide Redesign of Default Options Within the Electronic Health Record. *JAMA Intern Med*. 2016;176(6):847–848. PMID 27159011.
14. Meeker D, Knight TK, Friedberg MW, et al. Nudging Guideline-Concordant Antibiotic Prescribing: A Randomized Clinical Trial. *JAMA Intern Med*. 2014;174(3):425–431. PMID 24474434.
15. Patel MS, Volpp KG, Asch DA. Nudge Units to Improve the Delivery of Health Care. *N Engl J Med*. 2018;378(3):214–216. PMID 29342387.

Cite this document

ARKA Health. Three ways to change what a physician orders. Only two of them work. And what happens when you try the same thing on imaging. What we know about imaging; What we know about the mechanism; How the studies compare; What nobody knows; The

study we intend to run (v1.1, 16 August 2026, engine tcoc-1.0.0). <https://arkahealth.com/api/evidence/nudge-writeup>. Generated August 16, 2026. Retrieved August 16, 2026.

All patients, orders, and outcomes in illustrative examples are synthetic (ARKA-DEMO-001); no real patient data. Trial figures are measured from published sources. Version 1.1 · 16 August 2026.