

---

## TESTING RECORD

# OT Cost Analysis Testing Record

---

*Four testing waves across sixteen sub-agent runs in April 2026: 520 to 627 lines, 22 universal patches, all three statutory authorities exercised (10 USC 4021 research, 4022 prototype, 4022(f) production), six cost-sharing paths validated, zero reconciliation gaps at final regression.*

---

### Skill

ot-cost-analysis  
[github.com/1102tools/federal-contracting-skills](https://github.com/1102tools/federal-contracting-skills)

### Date

April 2026  
[1102tools.com](https://1102tools.com)

## The bottom line

Four testing waves across twelve cold sub-agent runs in April 2026 took the OT Cost Analysis skill from 520 to 627 lines through 22 universal patches (15 in the first structural wave, 7 in a refinement wave, 1 correctness fix). The skill reliably orchestrates BLS OEWS, GSA CALC+, and GSA Per Diem MCPs for labor benchmarking; handles all three statutory authorities (10 USC 4021 research, 4022 prototype, 4022(f) production follow-on); applies six distinct cost-sharing paths (A-NDC, A-via-sub, B-SB, C-traditional, D-competition, and research-inapplicable); supports mixed fixed-price / cost-type milestone structures with proper ceiling-based government obligation; correctly handles multi-performer and multi-MSA labor pools; and produces 7-sheet formula-driven workbooks where every user-adjustable assumption cascades through the entire build.

## Waves tested

Wave	Runs	Focus	Patches shipped	New structural gaps surfaced
1	4	Initial cold run, no MCPs wired	10 candidates identified	Most major gaps from this wave carried into Wave 2
2	4	Full MCP access, different paths	14 patches	7 new universal patterns
3	4	Patch regression on untested paths	7 patches (incl. 1 critical bug fix)	Confirmed patch bug: Performer Share didn't branch on Payment Type
4	4	Final regression on 7 fixes	1 critical correctness fix	4022(f) cost-share wording contradiction

## Authorities and paths exercised

#	Scenario	Authority	Performer path	Workflow
1	Group 1 sUAS autonomy, San Diego	10 USC 4021 prototype	NDC, (d)(1)(A)	A
2	Contested logistics AI planner, Arlington	10 USC 4021 prototype	Traditional, (d)(1)(C) 1/3 share	A
3	Cold-spray additive repair, concept	10 USC 4021 prototype	NDC, (d)(1)(A)	A+
4	AGV reasonableness check, Pittsburgh	10 USC 4021 prototype	Traditional, (d)(1)(D) competition	B
5	AUV port security, DIU consortium	10 USC 4021 prototype	SB, (d)(1)(B)	A
6	Quantum RF sensing, MIT Lincoln Lab	10 USC 4021 research	FFRDC, 4022(d) inapplicable	A
7	C-UAS interceptor LRIP, Albuquerque	10 USC 4022(f) production	NDC (inherited path), 100% gov	A
8	HMT platform, Leidos + NeuroForge	10 USC 4021 prototype	Traditional + NDC sub via (d)(1)(A)	A
9	Hypersonics mixed-type, Tucson	10 USC 4021 prototype	Traditional, (d)(1)(C) 1/3 share	A
10	Metamaterial RF, GTRI FFRDC	10 USC 4021 research	Traditional, inapplicable	A
11	ASV production 24 units, Mobile	10 USC 4022(f) production	Competition commitment (inherited), 100% gov	A
12	HMT platform Orlando+Boston	10 USC 4021 prototype	Traditional + NDC sub, multi-MSA	A

Plus the four Wave 4 regression tests (Northrop HEL reasonableness path C, Draper Boston concept via (d)(1)(A)(iii), Raytheon hypersonic seeker LRIP 100 units, NSTXL 3-performer space domain awareness).

## **Patches shipped in this skill**

**Wave 2 (14 patches, after MCPs wired and first round of structural gaps**

## surfaced)

Patch	Section affected	Trigger
CALC+ page_size=1 instruction (MCP rejects page_size=0)	Step 3 CALC+	Test 4 Wave 2 errored on verbatim skill instruction
10 USC 4021 vs 4022 authority gate before cost-share decision	Cost-Sharing Guidance	Test 6 Wave 2 caught silent mis-application of 4022(d) to 4021 research
10 USC 4022(f) production follow-on brought in-scope (was falsely routed to IGCE Builder)	"What This Skill Does NOT Cover"	Test 7 Wave 2 production follow-on would have mis-routed
Expanded SOC mapping for autonomy/ML/robotics/mechatronics/academic/production roles	Step 2a	Tests 1, 4, 5, 6, 8 all hit missing-SOC issue
Labor rate method: per-category canonical, blended only as quick pre-sol	Step 6 labor cost	Tests 2, 3, 4 picked different methods
Paid hours (2080) vs productive hours (1880) split explicit	Step 2d / Step 6	Tests 1, 2, 4 risked double-burden
Pre-solicitation mode promoted from edge case to first-class variant	Workflow A	Pre-sol is the most common use case
Workflow A+ Step 0 inlined TRL mapping + required performer location input	Step 0	Test 3 had to infer Huntsville from weak signals
Cost-type ceiling: B10 margin, Payment Type column, obligation at ceiling	Sheet 1 layout, Step 7	Tests 6 and 8 needed explicit treatment

Patch	Section affected	Trigger
Multi-performer prime+sub structure with Side column and Performer Structure block	Step 6	Test 8 required it
Multi-location labor: per-MSA BLS queries, per-row Location tags, no averaging	Step 2b	Test 8 hit it
Milestone duration vs PoP mismatch reconciliation rule	Step 1 validation, Step 6	Test 2 had \$515K swing on interpretation
Academic / FFRDC / UARC labor branch: shifted burden scenarios, grad RA institutional rates	Step 2 caveat	Tests 6 and 10 needed it
Production follow-on economics: labor mix, FY obligation profile, separate materials escalation	Step 2 production branch	Test 7 needed all three

### Wave 3 (7 patches, Wave 3 regression surfaced refinements)

Patch	Section affected	Trigger
<b>Critical: Performer Share IF branch on Payment Type</b>	Sheet 1 I column formula	Cost-type milestones + path (C) produced Sheet1-vs-Sheet5 reconciliation gap
Sheet 2 deterministic block placement via dict or defined names	Sheet 2 spec	Variable labor category counts broke fixed-offset assumptions
Assumption cell inline [\$B\$X] references	Sheet 1 assumption block	Model wrote wrong cell refs when label row and data row were confused
FY obligation mapping convention: obligate-at-start (default)	Production follow-on branch	Test 2 Wave 3 picked inconsistent conventions
Materials escalation time-basis: compound per milestone-start month from PoP start	Step 8 scenario analysis	Test 2 Wave 3 invented a formula in absence of prescription
Per diem FY fallback rule when target FY not yet published	Step 5 travel	Test 3 Wave 3 PoP started in unpublished FY
Learning curve operationalization: default 95% Crawford multiplier per lot for LRIP	Production follow-on branch	Test 3 Wave 3 per-unit cost was rising instead of falling

### Wave 4 (1 correctness fix)

Patch	Section affected	Trigger
10 USC 4022(f) cost-share wording contradiction resolved	Step 2 production branch	Authority Gate said "100% government funding"; Step 2 said "cost-share inherits." Opposite readings. Fixed to: path determination inherits, cost-share ratio does NOT propagate.



## What worked in every regression run

- Seven-sheet workbook schema: Summary, Milestone Detail, Scenarios, Labor Benchmarking, Cost-Sharing Detail + Funding Profile, Methodology / Price Reasonableness Memo, Raw Data
- Formula-driven cells throughout: changing any assumption cell (burden, cost-share ratio, escalation, ceiling margin, consortium fee) recalculates every dependent cell
- Blue-font convention for user-adjustable inputs vs black for formulas
- Pre-solicitation mode with conditional variance formulas: no proposed price required; formulas auto-activate when price is later entered
- Position flags against should-cost (below / within 10% / 10-25% / above 25%) applied consistently
- Methodology memo with 10 USC 4021/4022 authority citation, no FAR 15.404 dependence
- Correct handling of NDC, SB, Traditional, consortium fee, and cost-share paths in parallel

## What was not tested

- Hybrid FAR+OT arrangements (unusual but possible)
- Multi-award OT (government funding multiple performers on competing prototypes in parallel)
- Fixed-price-incentive-fee-style milestones (non-standard for OTs but seen in a few programs)
- OT with performance-based pay-for-outcome structures
- OCONUS performer with significant CONUS government integration travel
- Agreement structures above \$250M (skill tested up to ~\$60M)
- Multi-year options with government unilateral exercise

Users working in these contexts should expect to validate outputs more carefully.

## Testing methodology

Each wave consisted of four cold sub-agent runs. Each sub-agent read `SKILL.md` fresh with no conversation history and executed against a user prompt exactly as a `claude.ai` or Claude Code user would. All three MCPs (`bls-oews`, `gsa-calc`, `gsa-perdiem`) were live for Waves 2-4. Each agent produced a `.xlsx` workbook (or a described workbook if the environment did not support `.xlsx` generation) and a self-evaluation titled "Feedback for the evaluator" calling out patches that worked, patches that failed, and universal structural gaps. Findings were classified against a strict universal-only rule: only patches addressing structural patterns across multiple scenarios or clear statutory require-

ments were shipped; one-off model judgment failures and narrow domain fixes were deliberately skipped to prevent bloat.

Skill line count: 520 before Wave 1, 619 after Wave 3 patches, 627 after Wave 4. Ceiling remains 1,000.

## Known limitations

- openpyxl writes formula strings but no cached values. Workbooks recalculate on first open in Excel/Numbers. If distributed to an environment with formula recalc disabled, values will show #N/A until opened.
- Per diem FY rates beyond the published window require fallback to most recent published FY; the skill handles this but user should refresh per diem when new FY rates publish.
- CALC+ sparse-hit handling for niche defense roles (HEL, quantum, DEW) returns zero CALC+ results; skill falls back to BLS-only with methodology note; for these specialties, BLS medians understate actual market rates by 20-30%.

### Testing Methodology

Evaluator: James Jenrette (1102tools) and Claude Code Opus 4.7 (1M context window, max effort mode, Claude Max 20x subscription).

Worker model tested: Claude Opus 4.7 sub-agents with live MCP access to bls-oews, gsa-calc, gsa-perdiem.

Waves: 4 waves, 16 sub-agent runs total across the full testing program, 22 universal patches shipped (15 structural + 7 refinement + 1 correctness).

Date: April 2026.

Skill: ot-cost-analysis. Source: [github.com/1102tools/federal-contracting-skills](https://github.com/1102tools/federal-contracting-skills). License: MIT.