A Pre/Post Evaluation of a Two-Hour Stress-Competence Workshop in Women Physicians

Pilot data from DistressRx

Developed by Ali Novitsky, MD

Sample
42 women physicians
Matched pre/post survey respondents

Design
Within-subject pre/post
18 parallel 5-point Likert items

Intervention
Single 2-hour workshop
DistressRx framework

Summary

Background. Physician burnout and attrition are well-documented threats to clinical workforce sustainability. Most interventions reported in the literature are multi-session, multi-week, or contractual in scope. Brief single-session formats are comparatively understudied.

Methods. Forty-two women physicians completed parallel pre- and post-intervention surveys following DistressRx, a 2-hour workshop introducing the DistressRx framework, individual stress-subtype identification, and a primary coping strategy aligned to each subtype. Eighteen 5-point Likert items measured awareness, vocabulary, perceived skill, clinical presence, and leadership behaviors. Responses were matched by email. Paired comparisons used Wilcoxon signed-rank tests; effect sizes were calculated as Cohen's d on paired differences.

Findings. Across the 17 items measuring stress-competence, all 17 showed statistically significant improvement (14/17 at $p < .001$). Mean Cohen's d was 1.10 (median 1.09); 11/17 items met the large-effect threshold ($d \geq 0.8$). On a single burnout-symptom item, agreement increased from 3.07 to 4.26, consistent with a response-shift effect. Among the 17 participants who reported considering leaving their role at baseline, 11 (65%) post-intervention indicated that consistent stress-management support would reduce that intent.

Limitations stated at the outset

This is a within-subject pre/post pilot, not a randomized controlled trial. There is no control or wait-list comparison arm. Outcomes are self-reported. All participants were women physicians with prior coaching exposure, and findings may not generalize to other populations. Post-survey completion was voluntary; selection effects cannot be ruled out. Effect-size comparisons to published literature are descriptive and should not be interpreted as evidence of superiority to other interventions.

Sample characteristics

N = 42 women physicians who completed matched pre- and post-intervention surveys.

Years in clinical practice

Category	n	%
> 20	13	31.0
11–20	21	50.0
5–10	4	9.5
< 5	2	4.8
Not applicable	2	4.8

Role

Category	n	%
Practicing physician	26	61.9
Academic faculty	7	16.7
Physician leader / admin	5	11.9
Resident / fellow	1	2.4
Non-physician clinician	2	4.8

Practice setting

Category	n	%
Private practice	16	38.1
Hospital-employed	13	31.0
Academic medical center	9	21.4
Hybrid	2	4.8
Non-physician setting	1	2.4

Clinical FTE

Category	n	%
76–100%	30	71.4
51–75%	3	7.1
26–50%	4	9.5
0–25%	2	4.8
Not applicable	3	7.1

Baseline retention-intent

Q: "I am seriously considering leaving my current role in the next 6–12 months."

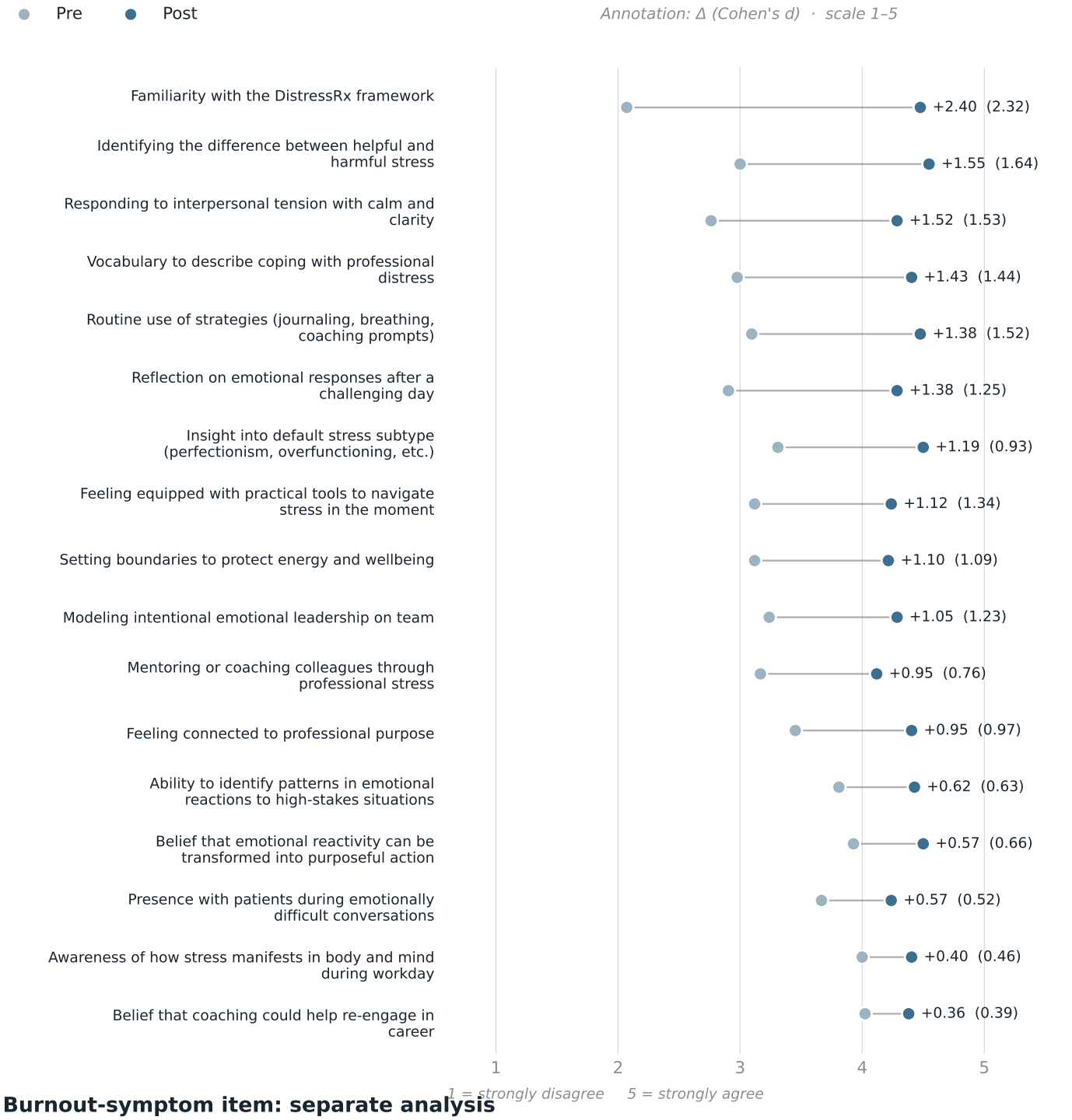
Yes	*Maybe*	*No*
6 / 42	**11** / 42	**25** / 42
14.3%	26.2%	59.5%

Endorsed reasons for considering departure (multi-select, all 42 participants)

Poor work-life integration (14) · Misalignment with values (12) · Burnout (9)
Different clinical structure (8) · Lack of support (7)

Item-level results

Paired pre- and post-intervention means for each of 17 stress-competence items, sorted by mean change. A separate item assessing burnout symptoms is reported below.



Burnout-symptom item: separate analysis

Q: "I currently feel symptoms of burnout (emotional exhaustion, cynicism, lack of accomplishment)."

Agreement with this item rose from a pre-mean of 3.07 to a post-mean of 4.26 (Wilcoxon $p < .001$). 21 of 42 participants (50%) shifted from a pre-rating below 4 to a post-rating of 4 or 5. This pattern is consistent with the response-shift phenomenon described in the wellness-intervention literature, in which exposure to new conceptual vocabulary alters the internal standard against which respondents evaluate themselves. The increase is therefore interpreted here as a change in symptom recognition rather than a change in underlying symptom severity, though both interpretations remain possible and cannot be distinguished without an external clinical measure.

Retention-intent findings

Pre- and post-survey items measured distinct but related constructs and are therefore reported separately.

Pre-intervention item

Q: "I am seriously considering leaving my current role in the next 6–12 months."

Response	*n*	*%*	*Bar*
Yes	6	14.3	
Maybe	11	26.2	
No	25	59.5	

Post-intervention item

Q: "If I had consistent stress-management support, I would be less likely to leave my current role in the next 6–12 months."

Response	*n*	*%*	*Bar*
Yes	30	71.4	
Maybe	8	19.0	
No	4	9.5	

Subgroup: participants at baseline retention risk

Among the 17 participants who answered "Yes" or "Maybe" to the pre-intervention item, post-intervention responses were:

Yes — stress support would reduce intent to leave	11/17	64.7%
Maybe	4/17	23.5%
No	2/17	11.8%

Interpretive note

These two items capture different constructs: the pre-survey item measures current intent to leave; the post-survey item measures a hypothetical response to consistent ongoing support. The findings indicate that participants identify stress-management support as a factor relevant to their retention decisions. They do not, however, demonstrate that the intervention itself reduced actual attrition; such a claim would require longitudinal follow-up of employment status.

Discussion

Principal findings

A single 2-hour workshop produced statistically significant self-reported improvement across 17 of 17 measured stress-competence dimensions in a sample of 42 women physicians, with a mean within-subject Cohen's d of 1.10. Items measuring vocabulary, awareness, and skill showed the largest changes; items measuring foundational beliefs and pre-existing clinical presence showed smaller changes, consistent with their already-elevated baseline means.

Interpretation of the burnout-symptom item

The increase in agreement with the burnout-symptom item from 3.07 to 4.26 is interpreted as a response-shift effect rather than a worsening of underlying burnout. This interpretation is consistent with the workshop's explicit teaching of burnout vocabulary and the parallel increase in the vocabulary item (Q12, d = 1.44). However, an actual increase in symptom severity cannot be ruled out with this design, and clinical follow-up would be needed to differentiate.

Context within the literature

Reported effect sizes for behavioral wellness interventions in physicians vary widely. Meta-analytic reviews of burnout-targeted interventions have generally found pooled effect sizes in the d = 0.2–0.4 range, with longer multi-session programs typically outperforming briefer formats. The effect sizes reported here exceed that benchmark; however, the comparison should be interpreted cautiously given differences in outcome measures, samples, and study designs. The current pilot uses a competence-and-awareness outcome set rather than a validated burnout instrument such as the Maslach Burnout Inventory, and direct comparison to studies using those instruments is not warranted.

Limitations

No control group; self-report only; selection effects from voluntary post-survey completion; participants had prior coaching exposure, which may have primed receptivity to the framework; sample of 42 limits subgroup analyses; women-only sample limits generalizability; outcome measures developed for this pilot are not externally validated.

Implications

The pattern of results suggests that brief single-session formats may have a useful role in physician stress education, particularly where multi-session programs are not operationally feasible. Replication with a control or wait-list comparison, validated burnout measures, and longitudinal follow-up of retention outcomes is warranted before stronger claims can be made.

Statistical methods: Paired comparisons by Wilcoxon signed-rank test (two-sided). Effect sizes computed as Cohen's d on paired differences (mean of difference / SD of difference). Conventional thresholds applied: small d ≥ 0.2, medium d ≥ 0.5, large d ≥ 0.8. Analyses conducted in Python using SciPy 1.x. No multiplicity correction applied; p-values reported as descriptive given the pilot nature of this analysis.