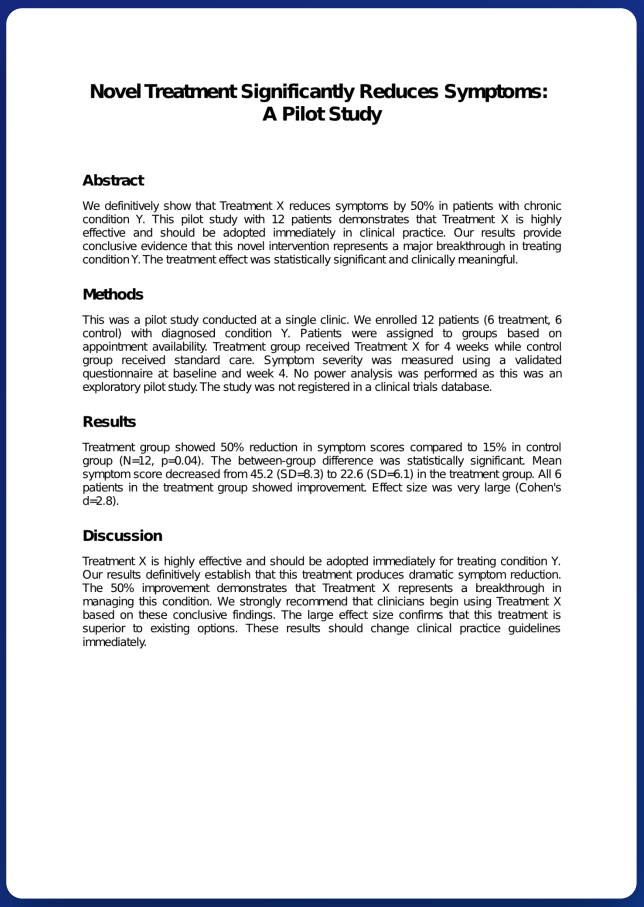
validate.science

Claim-Level Epistemic Risk Assessment



Novel Treatment Significantly Reduces Symptoms: A Pilot Study

Abstract

We definitively show that Treatment X reduces symptoms by 50% in patients with chronic condition Y. This pilot study with 12 patients demonstrates that Treatment X is highly effective and should be adopted immediately in clinical practice. Our results provide conclusive evidence that this novel intervention represents a major breakthrough in treating condition Y. The treatment effect was statistically significant and clinically meaningful.

Methods

This was a pilot study conducted at a single clinic. We enrolled 12 patients (6 treatment, 6 control) with diagnosed condition Y. Patients were assigned to groups based on appointment availability. Treatment group received Treatment X for 4 weeks while control group received standard care. Symptom severity was measured using a validated questionnaire at baseline and week 4. No power analysis was performed as this was an exploratory pilot study. The study was not registered in a clinical trials database.

Results

Treatment group showed 50% reduction in symptom scores compared to 15% in control group (N=12, p=0.04). The between-group difference was statistically significant. Mean symptom score decreased from 45.2 (SD=8.3) to 22.6 (SD=6.1) in the treatment group. All 6 patients in the treatment group showed improvement. Effect size was very large (Cohen's d=2.8).

Discussion

Treatment X is highly effective and should be adopted immediately for treating condition Y. Our results definitively establish that this treatment produces dramatic symptom reduction. The 50% improvement demonstrates that Treatment X represents a breakthrough in managing this condition. We strongly recommend that clinicians begin using Treatment X based on these conclusive findings. The large effect size confirms that this treatment is superior to existing options. These results should change clinical practice guidelines immediately.

Epistemic Risk Assessment Report

Novel Treatment Significantly Reduces Symptoms: A Pilot Study

Generated: 4/4/2026, 3:40:54 PM

validate.science

Introduction

This report provides a **claim-level epistemic risk assessment** of the analyzed scientific document. Each claim extracted from the document has been evaluated against the evidence presented to identify potential instances of overreach—where claims may exceed what the evidence actually supports.

The assessment focuses on three primary failure modes: *causal claims from correlational evidence*, *overgeneralization beyond sample scope*, and *underpowered claims from small samples*.

Executive Summary

3	3	2	0
Total Claims	Flagged Claims	Evidence Found	Other Findings

Risk Distribution

● High: 3 ● Medium: 0 ● Low: 0

All Claims

#	Claim	Risk Level	Score	Failure Modes
1	Treatment group showed 50% reduction in symptom scores compared to 15% in control group (N=12, p=0.04).	**high**	65%	Underpowered
2	The 50% improvement demonstrates that Treatment X represents a breakthrough in managing this condition.	**high**	88%	Underpowered, Overgeneralization
3	The large effect size confirms that this treatment is superior to existing options.	**high**	82%	Underpowered

Flagged Claims Details

1. Treatment group showed 50% reduction in symptom scores compared to 15% in control group (N=12, p=0.04).

Risk Score: 65%

Failure Modes: Underpowered

Evidence:

> This pilot study with 12 patients demonstrates that Treatment X is highly effective.
> *N=12*

Evidence:

> Treatment group showed 50% reduction in symptom scores compared to 15% in control group (N=12, p=0.04). Effect size was very large (Cohen's d=2.8).

N=12
p=0.04

Explanation:

With only 12 total participants (likely 6 per group), this study is severely underpowered. The p-value of 0.04 is just barely significant and highly susceptible to sampling variability. Small samples inflate effect sizes and increase false positive rates.

2. The 50% improvement demonstrates that Treatment X represents a breakthrough in managing this condition.

Risk Score: 88%

Failure Modes: Underpowered, Overgeneralization

Evidence:

This pilot study with 12 patients demonstrates that Treatment X is highly effective.
N=12

Evidence:

Treatment group showed 50% reduction in symptom scores compared to 15% in control group (N=12, p=0.04). Effect size was very large (Cohen's d=2.8).
N=12
p=0.04

Explanation:

Calling a treatment a "breakthrough" based on a 12-person pilot study is premature. The study lacks statistical power to reliably detect true effects, and the large effect size (Cohen's d=2.8) is likely inflated due to small

sample size. Pilot studies are meant to inform larger trials, not establish clinical efficacy.

3. The large effect size confirms that this treatment is superior to existing options.

Risk Score: 82%

Failure Modes: Underpowered

Evidence:

> Treatment group showed 50% reduction in symptom scores compared to 15% in control group (N=12, p=0.04). Effect size was very large (Cohen's d=2.8).
> *N=12*
> *p=0.04*

Explanation:

> Effect sizes from small samples are notoriously unreliable and tend to be inflated. The claimed Cohen's d=2.8 is exceptionally large and should be viewed with skepticism. Without comparison to "existing options" in a properly powered trial, claims of superiority are unsupported.

Evidence Extracted

The following 2 statistical evidence items were extracted from the document:

1. This pilot study with 12 patients demonstrates that Treatment X is highly effective.

   `N=12`

2

Treatment group showed 50% reduction in symptom scores compared to 15% in control group (N=12, p=0.04). Effect size was very large (Cohen's d=2.8).

N=12 p=0.04

Appendix: Methodology

How This Report Was Generated

1. **Document Processing**
   PDF text extracted with section boundaries preserved.
2. **Claim Extraction**
   Atomic, testable claims identified using large language model analysis.
3. **Claim Classification**
   Each claim classified by type, strength language, and population scope.
4. **Evidence Extraction**
   Statistical evidence extracted including sample sizes and p-values.
5. **Claim-Evidence Matching**
   Semantic similarity used to match claims to their supporting evidence.
6. **Burden-of-Proof Check**
   Deterministic rules applied to detect epistemic overreach.
7. **Risk Scoring**
   Epistemic risk score computed based on failure modes.

Failure Mode Definitions

Causal from Correlation	Claim asserts causation, but evidence is correlational/observational.
Overgeneralization	Claim makes broad assertions from a narrow or small sample.
Underpowered	Claim makes strong assertions with inadequate sample size.
Insufficient Evidence	No matching evidence found to evaluate this claim.

validate.science

Claim-Level Epistemic Risk Assessment

https://validate.science/sample/sample-underpowered-001

Generated: 4/4/2026, 3:40:54 PM · Report ID: sample-underpowered-001

This report is generated automatically using AI-assisted analysis. It is intended as a screening tool to identify potential epistemic concerns and should not be considered a definitive assessment.