

THE ECOSYSTEM OF EVIDENCE

Navigating the future
of the ecosystem of evidence

Artificial intelligence and digital transformation
in research and healthcare

11th EBHC International Conference
Taormina, 21st – 24th October 2026

#EBHC2026



MEDICAL UNIVERSITY OF WARSAW

30th 1996-2026 GIMBE EVIDENCE FOR HEALTH



How Uncertainty Visualization Shapes Interpretation of Comparative Results

A Systematic Review and Meta-Analysis

Łukasz Banaszek

Doctoral School
Department of Research Methodology
Medical University of Warsaw, Warsaw, Poland

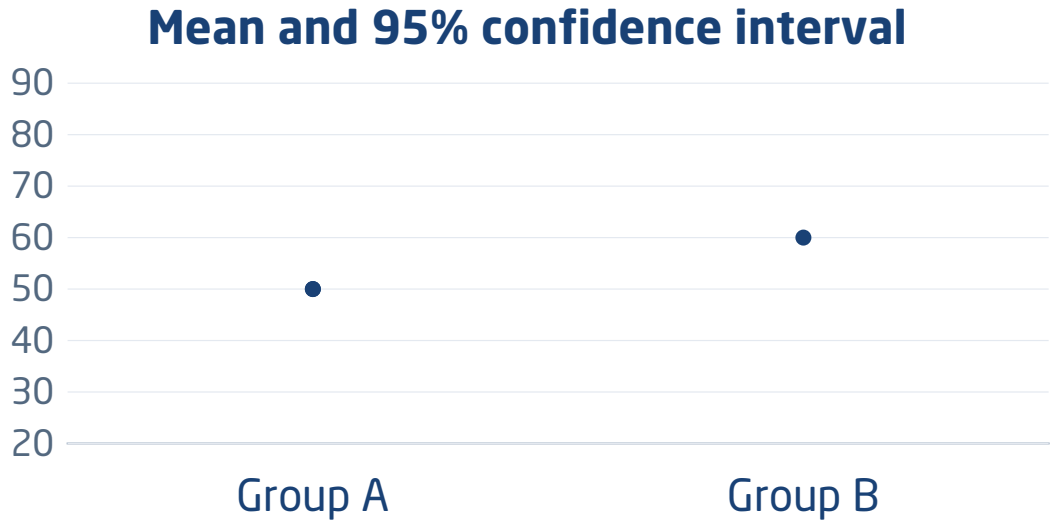
Aurelia Balcer, Karolina Pastwa, Albert Stachura, Paweł Włodar

Background

Inferential uncertainty (IU)

Standard errors (SE) and confidence intervals (CI)

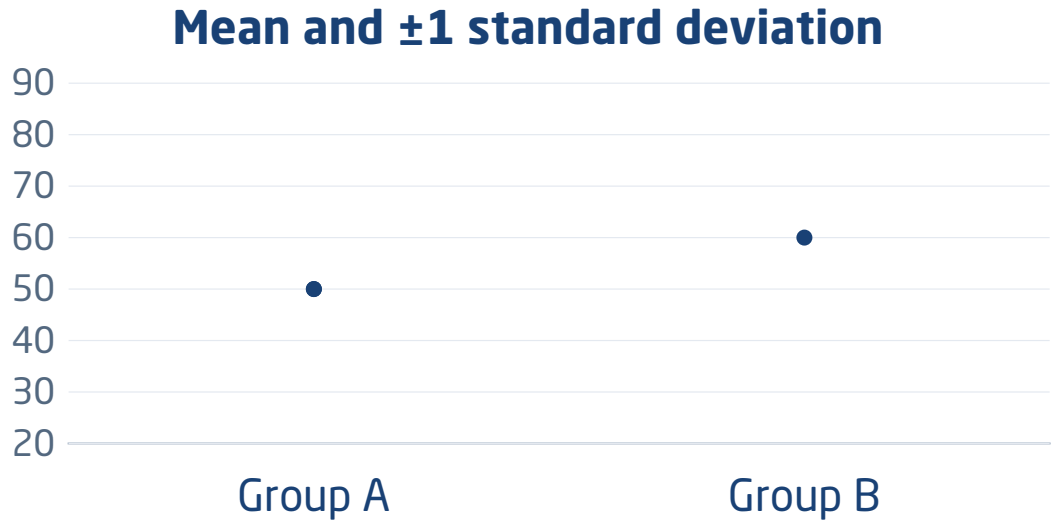
How precisely do we know
the average outcome?



Outcome variability (OV)

Examples: SD and individual outcome distributions

How much do individual
outcomes differ?



A narrow CI for the mean can coexist with widely varying individual outcomes

Illustrative data in arbitrary units. Same means, SDs and sample sizes in both panels (n = 100 per group).

Aims

How does uncertainty visualisation influence interpretation of comparative results?

Primary outcome

Perceived probability of superiority

Secondary outcomes

Valuation of an intervention, effect magnitude and certainty

Main comparison: inferential uncertainty versus outcome variability

Methods

Review component	Approach
Search	MEDLINE, Embase and Scopus No date restrictions, plus citation searching
Selection and extraction	Independent screening by three reviewers Data extraction in duplicate
Synthesis	Random-effects meta-analysis when outcomes were comparable Mean difference or Hedges' g
Appraisal	Revised JBI tool for quasi-experimental studies GRADE certainty of evidence

PRISMA 2020. PROSPERO CRD420251146953

Study selection and evidence base

257 database records
+ 13 citation records

→ **5** **studies included** →

2 studies in meta-analysis

14 superiority comparisons
12 willingness-to-pay comparisons
3 studies in narrative synthesis

Study	Participants	Task	Synthesis
Zhang 2023	Clinicians, data scientists and academic faculty	Treatment superiority and value	Meta-analysis
Hofman 2020	Lay participants	Intervention superiority and value	Meta-analysis
Belia 2005	Researchers	Statistical significance judgments	Narrative
Cumming 2004	Researchers	Predicting replication means	Narrative
Kale 2019	Lay participants	Identifying trends in noisy data	Narrative

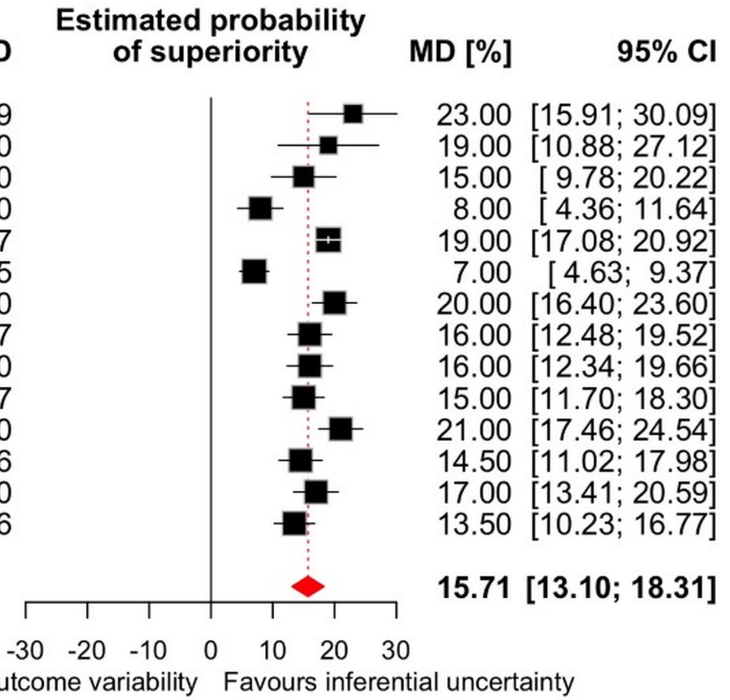
Perceived probability of superiority

MD +15.71 percentage points (95% CI 13.10 to 18.31)

I^2 = 87% Moderate certainty

Study ID	IU visualisation	OV visualisation	IU Total	IU Mean	IU SD	OV Total	OV Mean	OV SD	MD [%]	95% CI
Zhang 2023a	SE	SD	40	89.0	14.5	41	66.0	17.9	23.00	[15.91; 30.09]
Zhang 2023b	SE	SD	41	86.0	18.5	41	67.0	19.0	19.00	[10.88; 27.12]
Zhang 2023c	SE + mean	SE + mean + raw points	80	76.0	20.0	81	61.0	13.0	15.00	[9.78; 20.22]
Zhang 2023d	SE + mean	SE + mean + raw points	170	68.0	19.0	169	60.0	15.0	8.00	[4.36; 11.64]
Hofman 2020a	CI + matching text	PI + matching text	422	86.0	14.7	480	67.0	14.7	19.00	[17.08; 20.92]
Hofman 2020b	CI + extra text	PI + extra text	432	78.0	17.5	409	71.0	17.5	7.00	[4.63; 9.37]
Hofman 2020c	CI (small effect size)	PI (small effect size)	114	86.0	14.7	114	66.0	13.0	20.00	[16.40; 23.60]
Hofman 2020d	CI (large effect size)	PI (large effect size)	111	89.0	15.0	114	73.0	11.7	16.00	[12.48; 19.52]
Hofman 2020e	CI rescaled (small effect size)	PI (small effect size)	119	82.0	15.5	115	66.0	13.0	16.00	[12.34; 19.66]
Hofman 2020f	CI rescaled (large effect size)	PI (large effect size)	117	88.0	13.9	115	73.0	11.7	15.00	[11.70; 18.30]
Hofman 2020g	CI (small effect size)	HOP (small effect size)	114	86.0	14.7	106	65.0	12.0	21.00	[17.46; 24.54]
Hofman 2020h	CI (large effect size)	HOP (large effect size)	112	89.0	15.0	117	74.5	11.6	14.50	[11.02; 17.98]
Hofman 2020i	CI rescaled (small effect size)	HOP (small effect size)	120	82.0	15.5	106	65.0	12.0	17.00	[13.41; 20.59]
Hofman 2020j	CI rescaled (large effect size)	HOP (large effect size)	118	88.0	13.9	118	74.5	11.6	13.50	[10.23; 16.77]
Random effects model			**2110**			**2126**			**15.71**	**[13.10; 18.31]**

Heterogeneity: I^2 = 87.2%, τ^2 = 16.6951, p < 0.0001



2 studies (14 comparisons). Higher perceived superiority with IU displays.

Results

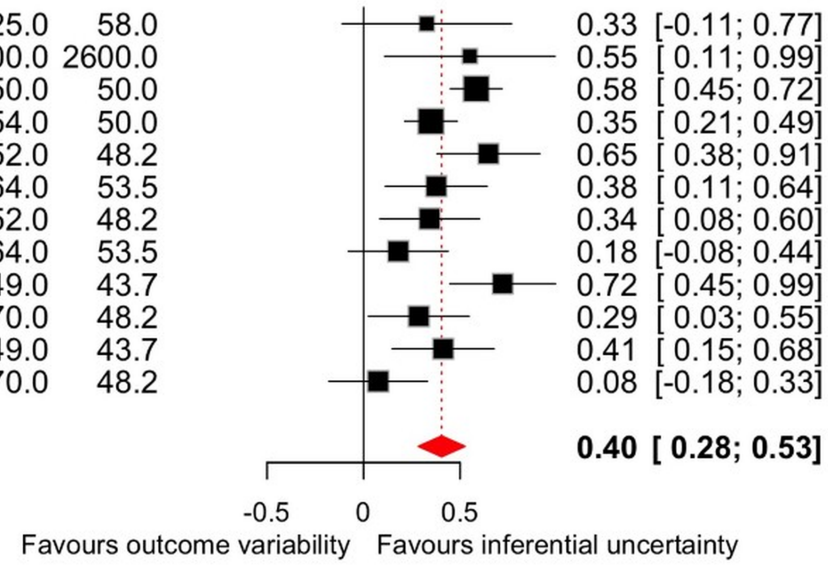
Willingness to pay

SMD (Hedges' g) = 0.40 (95% CI 0.28 to 0.53)

I^2 = 59% Low certainty

Study ID	IU visualisation	OV visualisation	IU Total	IU Mean	IU SD	OV Total	OV Mean	OV SD	SMD	95% CI
Zhang 2023a	SE	SD	40	45.0	63.0	41	25.0	58.0	0.33	[-0.11; 0.77]
Zhang 2023b	SE	SD	41	4000.0	2800.0	41	2500.0	2600.0	0.55	[0.11; 0.99]
Hofman 2020a	CI + matching text	PI + matching text	422	80.0	53.0	480	50.0	50.0	0.58	[0.45; 0.72]
Hofman 2020b	CI + extra text	PI + extra text	432	72.0	53.0	409	54.0	50.0	0.35	[0.21; 0.49]
Hofman 2020c	CI (small effect size)	PI (small effect size)	114	88.0	62.0	114	52.0	48.2	0.65	[0.38; 0.91]
Hofman 2020d	CI (large effect size)	PI (large effect size)	111	86.0	62.8	114	64.0	53.5	0.38	[0.11; 0.64]
Hofman 2020e	CI rescaled (small effect size)	PI (small effect size)	119	70.0	56.3	115	52.0	48.2	0.34	[0.08; 0.60]
Hofman 2020f	CI rescaled (large effect size)	PI (large effect size)	117	74.0	57.5	115	64.0	53.5	0.18	[-0.08; 0.44]
Hofman 2020g	CI (small effect size)	HOP (small effect size)	114	88.0	62.0	106	49.0	43.7	0.72	[0.45; 0.99]
Hofman 2020h	CI (large effect size)	HOP (large effect size)	112	86.0	62.8	117	70.0	48.2	0.29	[0.03; 0.55]
Hofman 2020i	CI rescaled (small effect size)	HOP (small effect size)	120	70.0	56.3	106	49.0	43.7	0.41	[0.15; 0.68]
Hofman 2020j	CI rescaled (large effect size)	HOP (large effect size)	118	74.0	57.5	118	70.0	48.2	0.08	[-0.18; 0.33]
Random effects model			**1860**			**1876**			**0.40**	**[0.28; 0.53]**

Heterogeneity: I^2 = 58.5%, τ^2 = 0.0205, p = 0.0054



2 studies (12 comparisons). Higher valuations with IU in hypothetical tasks.

Narrative findings

Belia 2005
Significance judgments

22% of significance judgments met the study's accuracy criterion

Cumming 2004
Replication expectations

Researchers underestimated the variability of future replication means

Kale 2019
Trend discrimination

Hypothetical outcome plots improved trend judgments versus static error bars

Distinct tasks and outcomes, synthesised narratively

Limits

The direction is consistent, but the evidence is limited

Outcome	GRADE certainty	Reason for downgrading
Perceived superiority	Moderate	Inconsistency in effect magnitude (I^2 = 87%)
Willingness to pay	Low	Inconsistency (I^2 = 59%) and indirectness

Primary outcome: 14 comparisons from only two studies

Effects varied across formats, tasks and participant groups

Impact on real clinical decisions remains uncertain

JBI appraisal: 2 studies at low risk, 3 with some concerns. No study at high risk.

Conclusions

Displaying the same data differently can change perceived superiority and value

Inferential uncertainty displays increased perceived superiority and valuation relative to outcome variability

Readers should distinguish precision of an average from variability in individual outcomes

Whether combined displays improve clinical decisions requires direct testing

No specific funding. The authors declare no competing interests.