

**Blue
Adapt**

Testing Hypothetical Bias Countermeasures in Discrete Choice Experiments

Evidence from an Embedded Plausibility Test

Wojciech Zawadzki¹, Mikołaj Czajkowski¹, Milan Ščasný²

¹ University of Warsaw, Poland ² Charles University, Czech Republic



BlueAdapt is funded by the European Union's Horizon Europe research and innovation programme under grant agreement No 101057764 and by the UKRI/HM Government.



Co-funded by
the European Union



UK Research
and Innovation



climate-health
cluster

Hypothetical bias: a long-standing threat to DCE credibility

A novel plausibility test to benchmark hypothetical bias countermeasures

- **Stated preference / DCE methods are the workhorse for non-market valuation**
 - But respondents choose in hypothetical settings, with no real consequences
- **Hypothetical bias: stated WTP/WTA diverges from real behavior**
 - A long-standing threat to the credibility of DCE-based welfare estimates
- **Common ex-ante countermeasures show mixed, inconclusive evidence**
 - Cheap talk, solemn oath, opt-out reminders, attribute-order changes
- **This study: a randomized DCE on coastal bathing-site choice (n=3,312)**
 - Travel distance as a salient, non-monetary numéraire
 - Area of interest: Gulf of Gdańsk & Vistula Lagoon, Poland
- **Key innovation: an embedded 13th "plausibility test" choice task**
 - Flags implausible high-distance choices → used to benchmark 4 hypothetical bias countermeasures

Study design: an embedded plausibility test

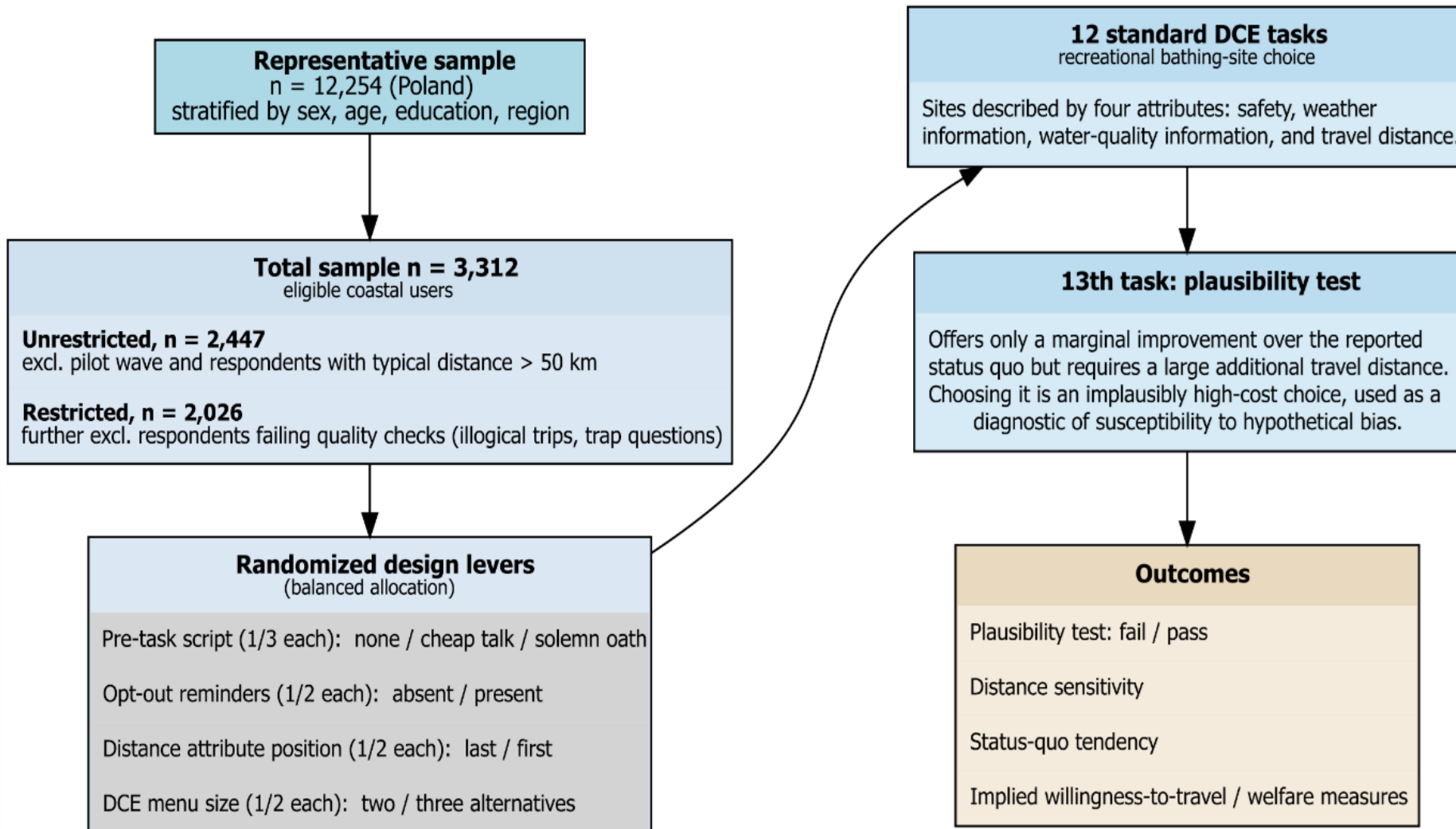
Recreational users of Gulf of Gdańsk & Vistula Lagoon coastal amenities (n=3,312)



www.blueadapt.eu



@BlueAdaptEU



The DCE: an example choice task

12 choice tasks drawn from the experimental design, plus 1 diagnostic task

Which of these locations (bathing sites) would you go to for recreation or outdoor water recreation?			
	Location A (status quo)	Location B	Location C
Designated bathing site	Based on the respondent's previous answers	Yes	Yes – lifeguarded
Information on temperature, wind strength, and the state of the sea		Yes	No
Information on water quality		Yes - once every 1 month	Yes - once every 2 weeks
Distance (in km)		Distance (SQ) + 25% km	Distance (SQ) + 50% km
Your choice:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

• Attributes & levels

- Designated bathing site: none / designated / lifeguarded
- Weather information board: no / yes
- Water-quality monitoring: none, every 2 weeks, weekly, every 2 days, daily
- Distance (numéraire): pivoted at -50%, -25%, 0%, +25%, +50%, +100% of the respondent's usual trip

The DCE: an embedded plausibility test

13th choice task – artificial indicator of hypothetical bias



www.blueadapt.eu



@BlueAdaptEU

Which of these locations (bathing sites) would you go to for recreation or outdoor water recreation?			
	Location A (status quo)	Location B	Location C
Designated bathing site	Based on the respondent's previous answers	The same as in SQ	The same as in SQ
Information on temperature, wind strength, and the state of the sea		The same as in SQ	The same as in SQ
Information on water quality		The same as in SQ	One level above SQ
Distance (in km)		Distance (SQ) - 50% km	Distance (SQ) + 50 km
Your choice:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- **Plausibility test**

- Designated to flag "suspicious" responses (unlikely to be welfare-consistent; potential indicator of hypothetical bias)
- Can't distinguish genuine preferences of respondents who just like to travel quite far
- In the context of travel within the Gulf of Gdańsk and the Vistula Lagoon region, visitors typically choose accommodation close to the coastline and tend to frequent the nearest bathing site, occasionally visiting slightly more distant alternatives. Traveling an additional 50 km along the Polish coast would take approximately one hour and would more plausibly incentivize individuals to select accommodation closer to the desired site rather than to undertake such travel repeatedly during a multiple-day stay.
- While imperfect, this plausibility test diagnostic is well-suited for comparative analysis across treatments, particularly when examining differences in the share of respondents selecting distance-insensitive alternatives.

Treatment A: pre-task scripts

Cheap talk vs. solemn oath vs. no script (baseline, TA0)



www.blueadapt.eu



@BlueAdaptEU

- **Cheap talk script (TA1)**

- Explicitly tells respondents that choices are hypothetical but should be answered as if consequential
- Highlights that respondents in similar studies tend to overstate their willingness to travel
- Asks them to weigh whether extra amenities justify the extra distance

- **Solemn oath script (TA2)**

- Does not mention hypothetical bias explicitly
- Asks respondents to personally promise to answer as they would behave in a real setting
- Respondents unwilling to promise are excluded from the analysis

TA0.	TA1.	TA2.
No pre-task info	Cheap-talk script	Solemn oath script
<i>Are you ready? Let's begin!</i>	<p><i>The choice situations (we will present to you shortly) are, of course, hypothetical. No one will force you to undertake a particular trip even if you choose it. In many studies of this type, it has been observed that respondents often choose places they would not actually visit because they are too far away. We call this the hypothetical bias. In this regard, we have a request for you.</i></p> <p><i>For each choice situation, consider whether the additional amenities at the beach are worth the extra distance you would have to travel. Remember, you can always select "Location A", which corresponds to your typical trip and does not involve any extra distance.</i></p> <p><i>We ask that you choose beaches exactly as if you were really going to face the consequences of your decision: that is, to actually go to a beach within a given distance</i></p> <p><i>Are you ready? Let's begin!</i></p>	<p><i>In this survey, it is crucial for us to understand the decisions of individuals as if they were really choosing a bathing site for recreation. In this regard, we ask you to make the following declaration.</i></p> <p><i>I promise that in the following questions, I will answer according to how I would behave in a real situation - that is, I will choose only those bathing sites to which I would actually go, considering the distance to them.</i></p> <ul style="list-style-type: none"><input type="checkbox"/> <i>I promise</i><input type="checkbox"/> <i>No promises</i> <p><i>Are you ready? Let's begin!</i></p>



Treatments B-D: opt-out, attribute order, menu size

Low-cost design levers embedded directly in the choice tasks

- **Opt-out reminder (B) & attribute order (C)**

- Opt-out: a one-line reminder under the choice question – “*NOTE: Choose only the locations (bathing sites) you would actually go to, considering the distance to them.*”
- Attribute order: distance listed last (TC0, conventional) vs. first (TC1, to test salience)

- **Menu size (D) & the 13th task**

- Menu size: 2 alternatives (status quo + 1) vs. 3 alternatives (status quo + 2)
- 13th task: marginal attribute improvement + 50 km detour; choosing it = “fail”, the plausibility-test diagnostic

Analytical approach

Two complementary models: failure diagnostics and preferences

- **Binary logit + average marginal effects**

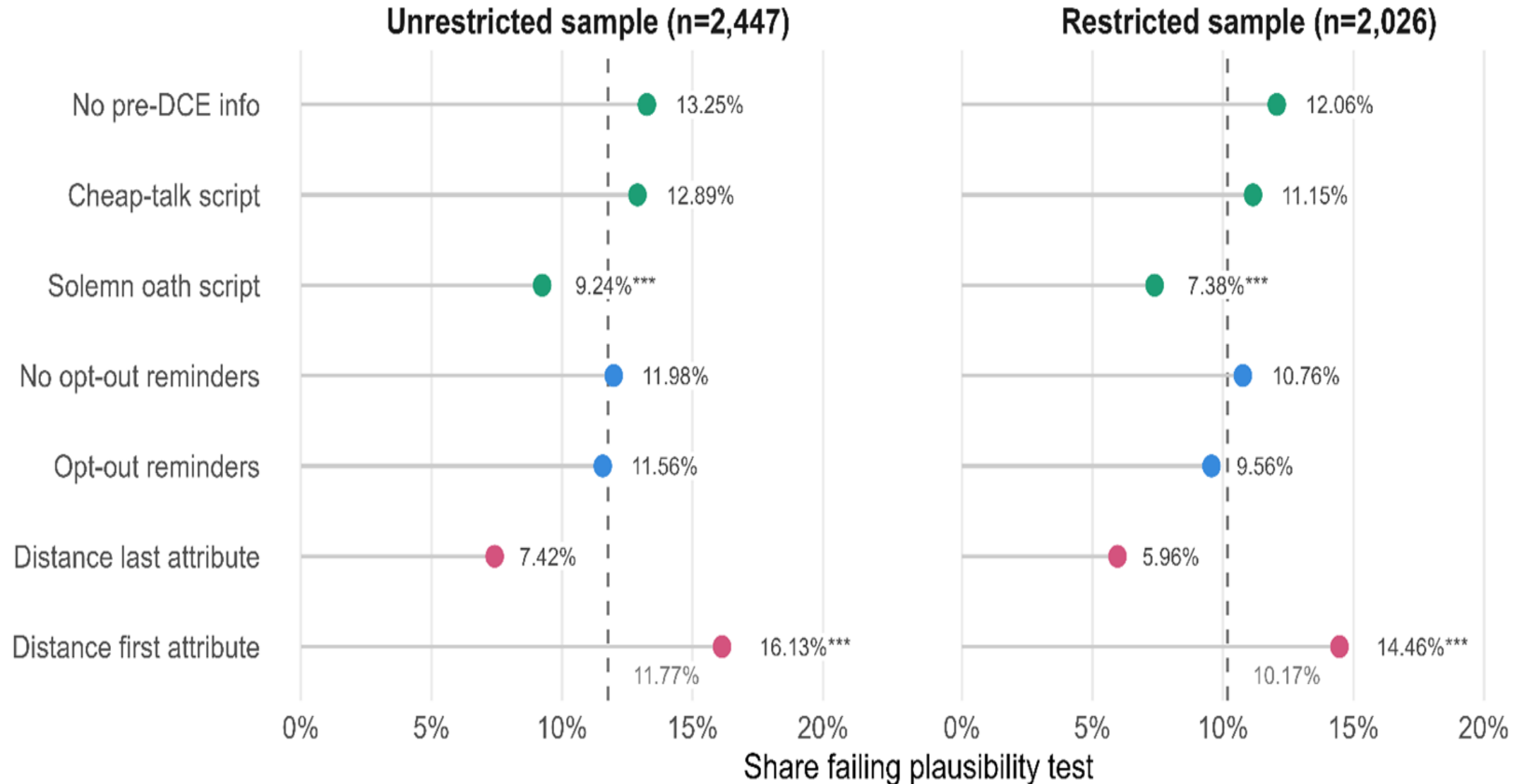
- Outcome: fail vs. pass the plausibility test
- Covariates: 5 treatments, all two-way interactions, plus socio-demographic controls
- Reported as average marginal effects (AMEs) with heteroskedasticity-robust standard errors
- Holm-adjusted p-values control for multiple comparisons across pre-specified hypotheses

- **Mixed logit in WTP-space (WTT-space)**

- Correlated random parameters: Normal for both non-cost attributes and distance
- 10,000 scrambled Sobol draws (Owen + Faure-Tezuka scrambling) for simulated maximum likelihood
- Treatment interactions enter the means, to estimate how countermeasures shift WTT
- Estimated on 3 samples (unrestricted / restricted / excl. cs13 failures), pooled and by menu size

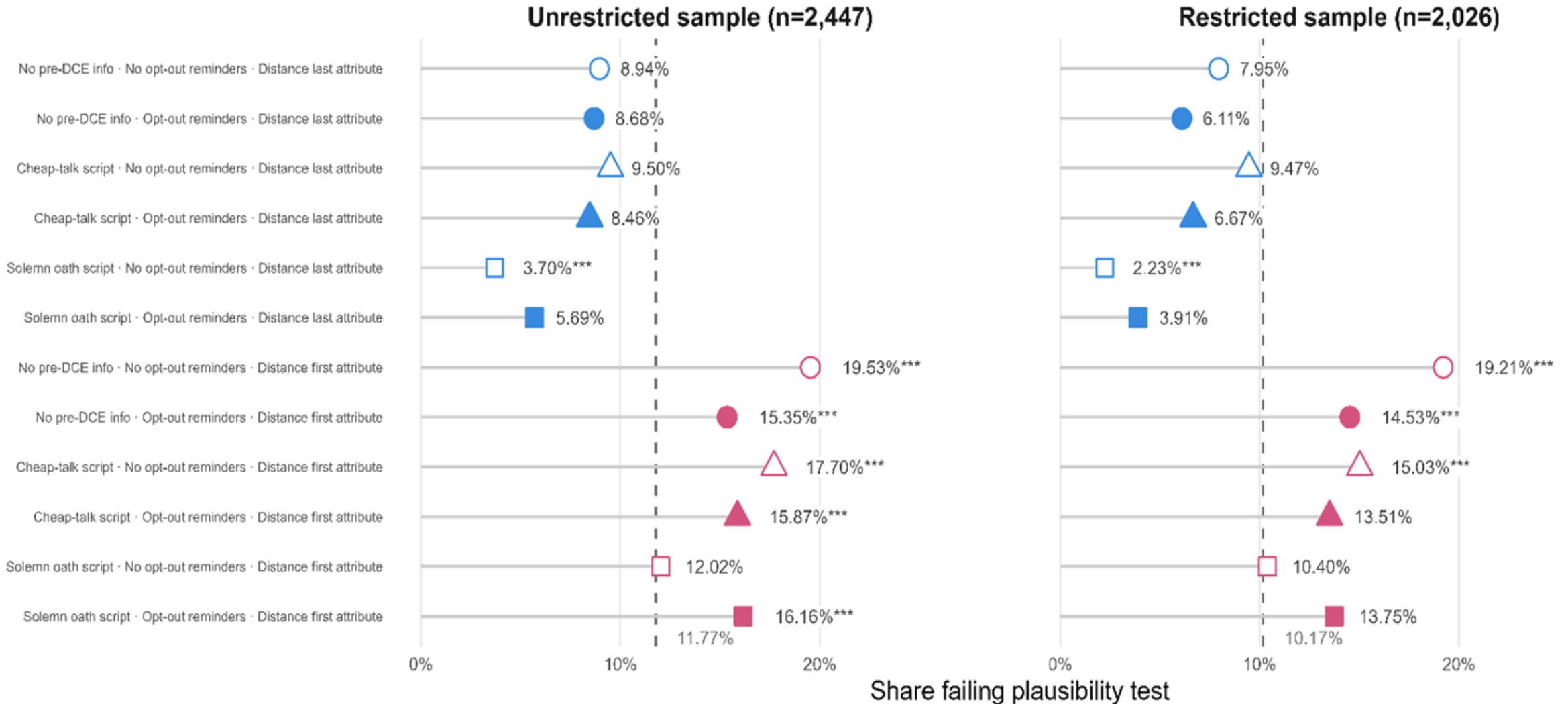
Results: plausibility-test failure rates by treatment

Solemn oath cuts failures – attribute order backfires



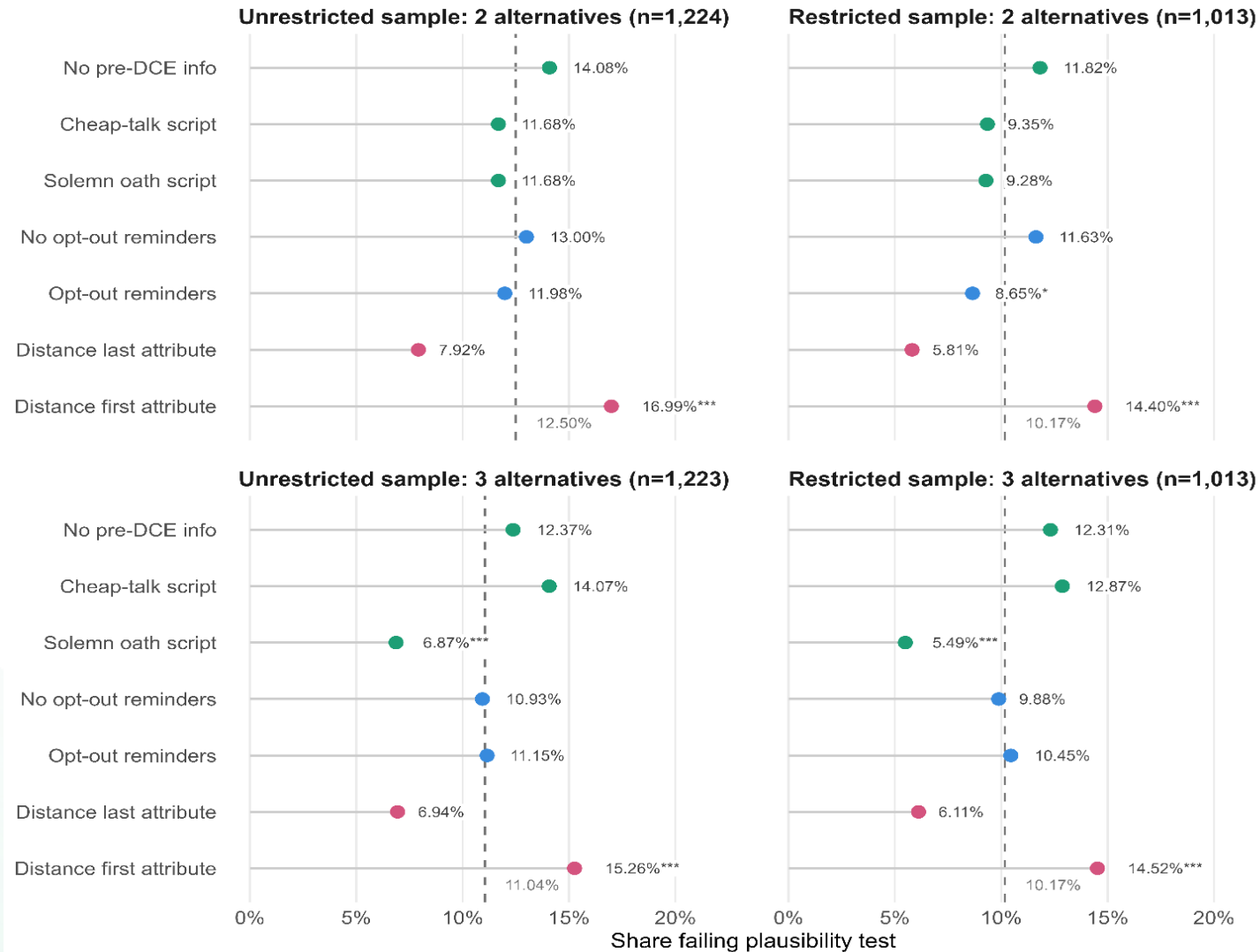
Results: failure rates by treatment combination

Effects hold up across every treatment combination



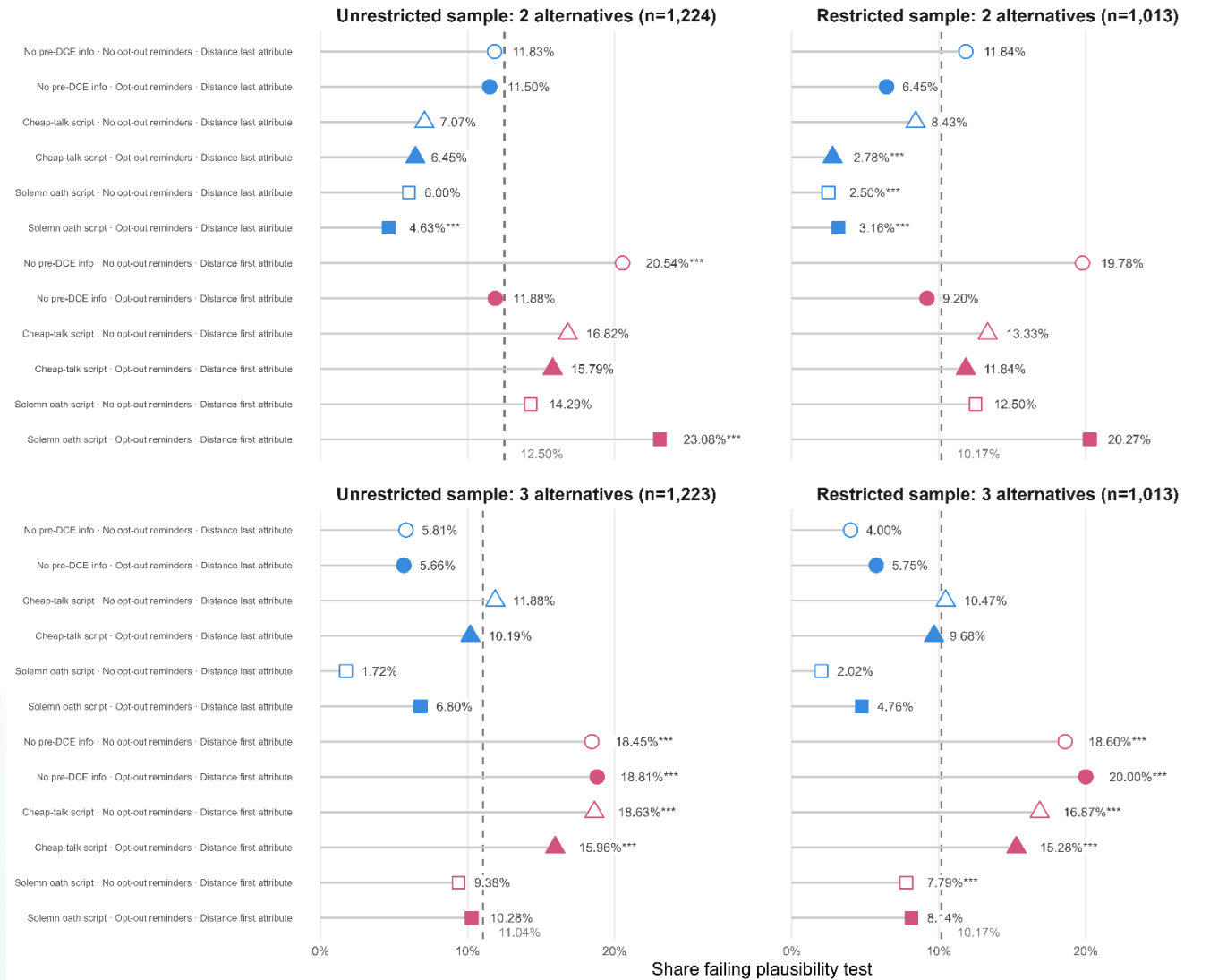
Robust across menu sizes

Failure rates by treatment, split by DCE menu size



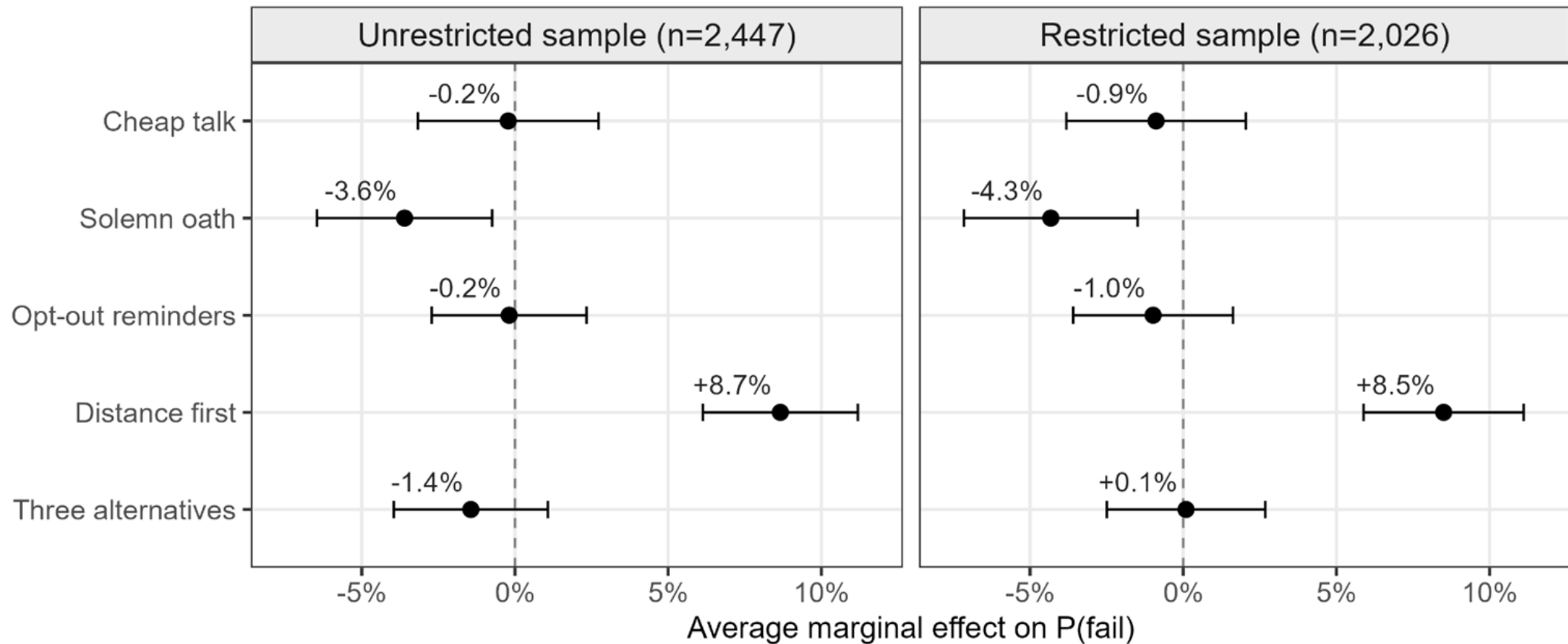
Distance-first still backfires within each menu size

Failure rates by treatment combination, split by menu size



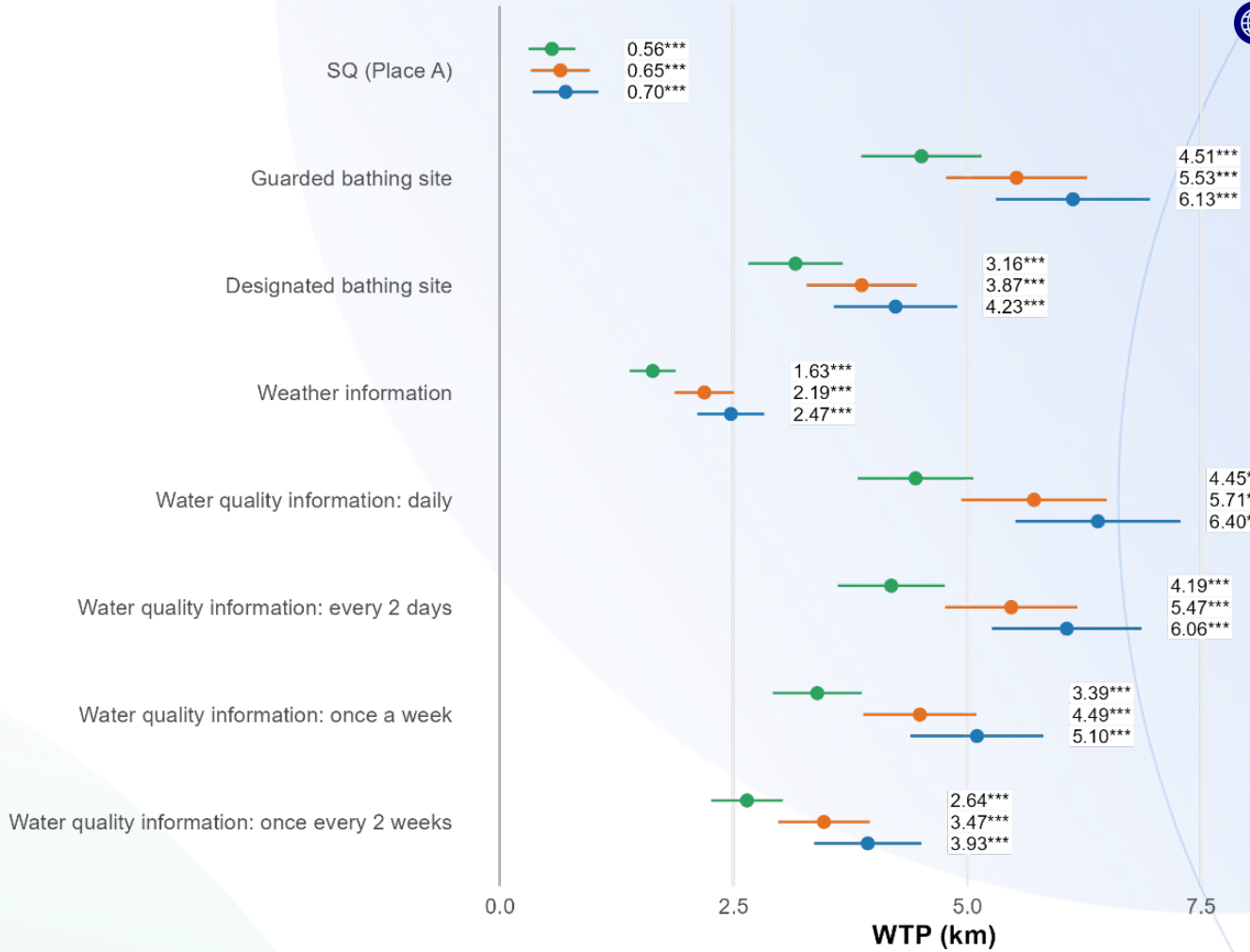
Solemn oath and attribute order dominate the regression too

Binary logit average marginal effects on P(fail), with controls



Willingness to travel: the headline estimates

WTT estimates fall as the sample becomes more restrictive

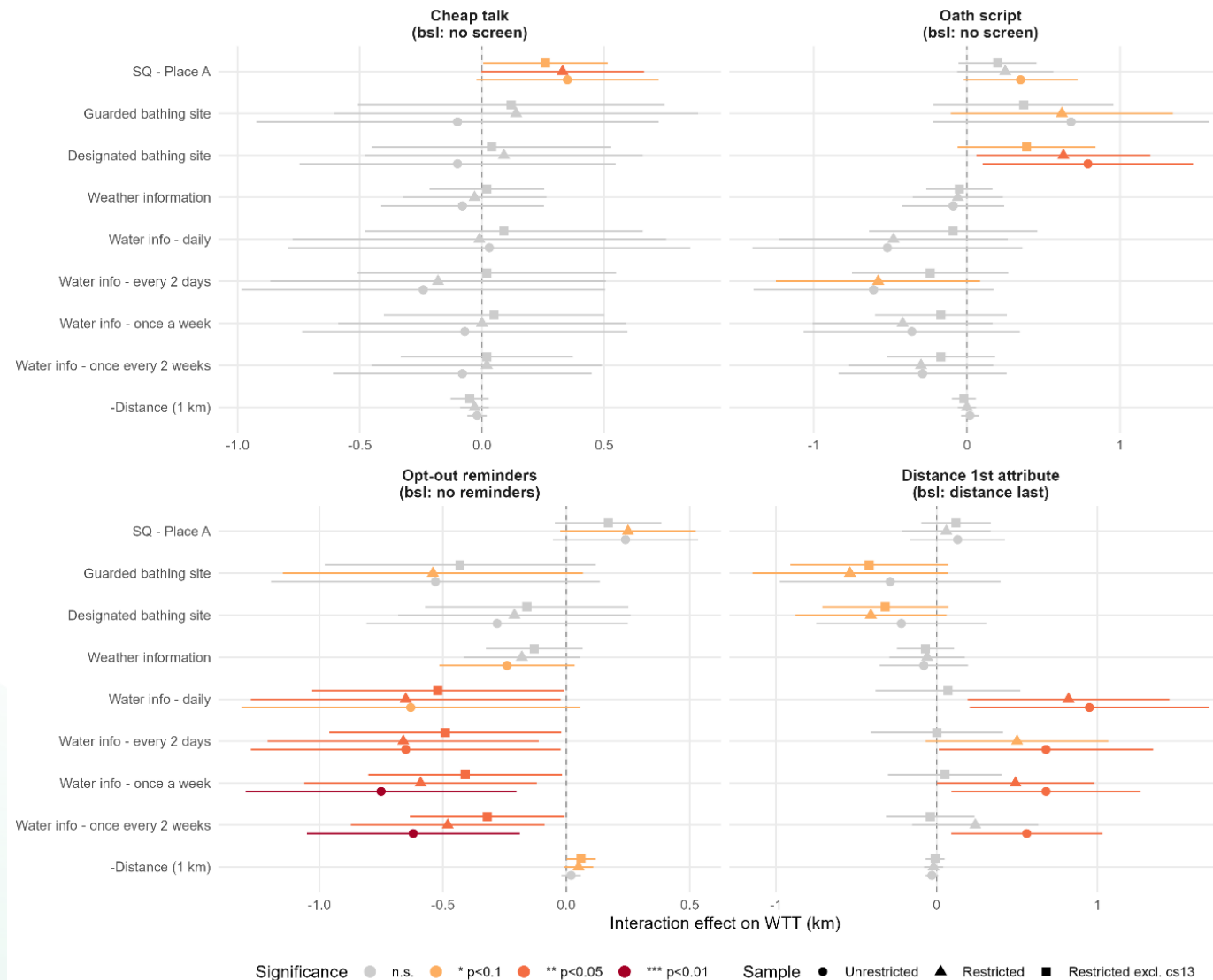


Subsample ● Unrestricted ● Restricted ● Restricted excl. implausible cs13 choice



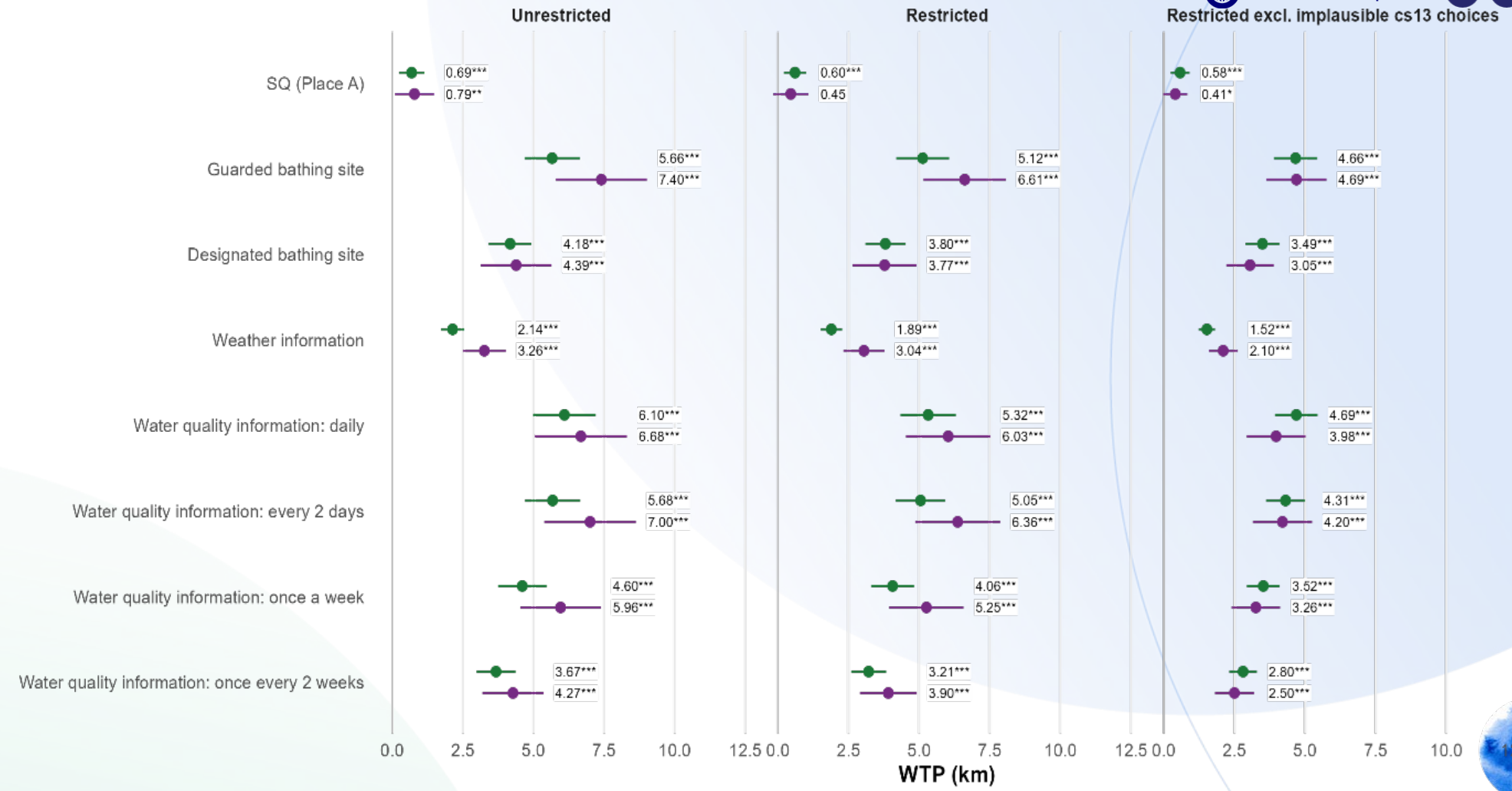
Countermeasures reshape preferences

Interaction effects of each treatment on WTT (WTP-space)



WTT patterns are broadly consistent across menu sizes

Willingness-to-travel estimates split by 2 vs. 3 alternative DCE



Menu size — 2 alt — 3 alt

Treatment effects on WTT also vary by menu size

Interaction effects on WTT, split by 2 vs. 3 alternative DCE



Discussion: cost salience and attribute order

Placing distance first backfires – why?

- **Possible mechanisms**

- Attention depletion / anchoring: distance is registered first, then discounted as more attractive attributes follow
- Last-row heuristic: the attribute encountered right before the choice is most behaviorally salient
- Attribute-order framing: leading with distance can resemble an abstract comparison, not a destination choice
- Interaction with attractive improvements: safety/quality gains may outweigh an early-registered distance cost

- **Implications**

- Conventional practice (numéraire last) looks behaviorally important, not just a stylistic convention
- Placing distance first also shifts estimated preferences – a sign the shift may be artificial, not genuine
- Attribute-order effects create scope for researcher-induced manipulation; keeping cost last limits this risk

Discussion: which countermeasures actually work?

Solemn oath outperforms cheap talk and opt-out reminders

- **Solemn oath**

- Most consistent effect; corroborates Carlsson et al. (2013) and De-Magistris & Pascucci (2014)
- Relies on honesty priming / moral commitment rather than economic incentives – raises ethical questions
- Performs better without opt-out reminders; effect weaker in the 2-alternative DCE

- **Cheap talk & opt-out reminders**

- Cheap talk: weakest, most context-dependent; counterproductive in the 3-alt DCE
- Opt-out reminders: modest, mostly insignificant effect on failure rates, but a more intrusive effect on preferences
- Both explicitly reference or nudge around hypothetical bias – unclear if this changes real preference revelation

Discussion: menu size and the artifact question

Fewer implausible choices, but also reshaped stated preferences

- **DCE menu size**

- Binary choice resembles an advisory referendum (Carson & Groves, 2007); richer menus aid trade-offs but may invite heuristics
- 3-alt DCE has a lower baseline failure rate, but the plausibility test is not directly comparable across formats
- Countermeasure performance is broadly consistent, though the oath is less effective in the 2-alt format alone

- **Genuine preferences or design artifact?**

- Mixed logit shows treatments shift distance sensitivity and status-quo choice, not only pass/fail rates
- Lower stated WTP/WTT is not, by itself, evidence that hypothetical bias has been mitigated
- *Open question: does a countermeasure elicit more genuine preferences, or just change the decision process?*

Conclusions & practical recommendations

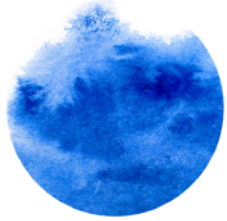
Benchmark countermeasures against behavioral diagnostics, not welfare estimates alone

• Three headline findings

- Solemn oath most consistently reduces plausibility-test failures (to ~2-4%); cheap talk & opt-out are weaker, context-dependent
- Placing distance first increases failures (up to 23%) – keep the numéraire last
- Countermeasures reshape stated preferences and welfare estimates, not just failure rates

• Practical recommendations

- Keep the cost / numéraire attribute last on the choice card
- Consider a solemn oath script, but weigh its ethical / moral-commitment concerns transparently
- Evaluate hypothetical-bias tools with an independent behavioral diagnostic, not lower valuations alone
- Future work: monetary DCE contexts, refined diagnostics, external validation against real behavior



**Blue
Adapt**

Thank You

Wojciech Zawadzki | wm.zawadzki@uw.edu.pl

Find out more at www.blueadapt.eu

Follow us   @BlueAdaptEU



Co-funded by
the European Union



UK Research
and Innovation



climate-health
cluster

BlueAdapt is funded by the European Union's Horizon Europe research and innovation programme under grant agreement No 101057764 and by the UKRI/HM Government.