Australian Government
Department of the Prime Minister and Cabinet

10 YEARS OF BETA

Meta-analysis of our behavioural trials

November 2026

Contents

Technical appendix 3

References 18

Technical appendix

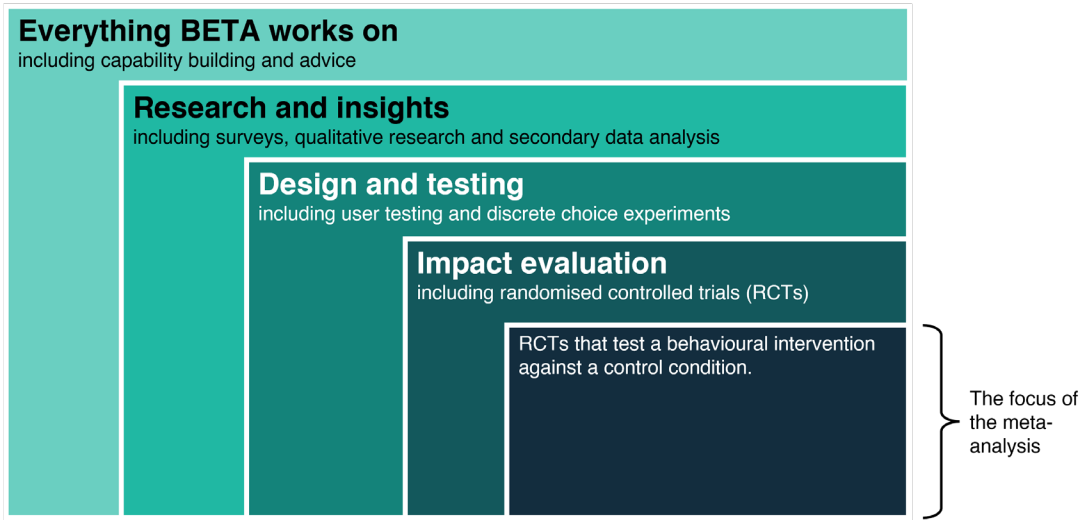
Overview

For the last decade, the Behavioural Economics Team of the Australian Government (BETA) has helped the Australian Public Service (APS) and policymakers design and test evidence-informed solutions, evaluate how interventions are working in the real world, and increase the use of Randomised Controlled Trials (RCTs).

This meta-analysis synthesises evidence from BETA's RCTs that compared behavioural interventions with an eligible control condition. The final analysis included 148 effect estimates from 52 trials across 40 projects and found a small positive average effect. This effect corresponds approximately to an additional 10 to 20 people out of every 1,000 taking the action compared with what would have happened without the intervention. The estimate was derived from the primary meta-analysis result (d = 0.07) by translating the average standardised effect size into an approximate absolute change in behaviour using benchmark assumptions about baseline behaviour rates. These small but important improvements accumulate into significant societal benefits.

This meta-analysis includes our published trials where BETA's behavioural interventions were relative to a control. As such, other impactful work completed by BETA is outside the scope of this analysis (Figure 1). This includes advice, qualitative or exploratory work, as well as trials or trial arms that compared interventions with each other rather than a control condition. As such, the meta-analysis provides a robust estimate of the average impact of eligible behavioural interventions relative to a control condition but is not a comprehensive account of all BETA's work.

Figure 1 The subset of trials included in the meta-analysis



A 3-level random-effects model was used because some trials contributed multiple eligible effect estimates. We found a small positive effect across BETA's portfolio. This finding was robust to alternative modelling approaches and was supported by the primary frequentist and Bayesian meta-analytic models. Moderator analyses found no evidence that intervention or trial characteristics systematically explained variation in effect sizes. While the average effect was positive, there was substantial variation in effects across trials and projects. This suggests that behavioural interventions cannot be assumed to work equally well across all settings and emphasises the need to test solutions with people within the specific policy context as a cost-effective way to deliver policy.

Methods

Study scope and eligibility

The evidence base for the meta-analysis comprised completed randomised controlled trials conducted by BETA over the past 10 years (2016-2026).

An outcome was eligible for quantitative synthesis if it:

- came from a BETA randomised controlled trial
- measured an eligible behavioural outcome
- compared an intervention with a stable control or other eligible baseline
- provided sufficient information to calculate an effect size and its sampling variance.

Eligibility and selection decisions were applied separately to each outcome within a trial. A trial could therefore contribute estimates for multiple distinct behavioural outcomes.

Selection of outcomes and contrasts

Trial effect estimates were selected for the meta-analysis to maintain a consistent intervention–control comparison and avoid biases by counting the same evidence more than once.

Comparators

The primary analysis included comparisons between an intervention and a stable baseline or control condition, such as, business-as-usual or an inert attention control (that is, where the treatment group received some contact, such as general information or check-in calls, but without the active intervention).

Where a trial did not include a conventional control condition, another trial arm could be used as the comparator only if it:

- provided a stable baseline
- contained no active behavioural content
- was not treated as an intervention elsewhere in the trial.

Treatment–treatment comparisons were excluded. Comparisons between 2 active interventions estimate their relative effectiveness rather than the effect of an intervention against a stable baseline and were therefore outside the scope of the primary analysis.

Multiple outcomes and timepoints

A trial could contribute to the meta-analysis separate effect estimates for distinct eligible behavioural outcomes, but only one estimate was retained for each outcome. Repeated measurements, alternative specifications and multiple representations of the same outcome were not treated as independent evidence.

Where an outcome was measured at multiple timepoints, the pre-specified primary endpoint was used. If this was unclear, the measurement taken at the end of the intervention period was selected.

Multi-arm trials

In trials with multiple intervention arms, each eligible intervention arm could be compared with the common control group where the comparisons represented pre-specified primary hypotheses. This meant that some control groups contributed to more than one effect estimate. The resulting effect estimates were treated as dependent observations within the same trial and modelled using the 3-level meta-analytic framework.

Pooled and arm-specific comparisons were not both included for the same outcome.

- Where the pooled comparison represented the primary research question, the pooled estimate was retained and the arm-specific estimates were excluded.
- Otherwise, the pre-specified arm-specific comparisons were retained.

Factorial designs

For fully randomised factorial trials that tested 2 or more interventions including whether the interventions work differently when combined, the analysis included the main effect corresponding to the behavioural intervention of interest, averaged across levels of the other randomised factor. Interactions were recorded but were not included as separate effect estimate in the meta-analysis unless comparable interaction effects were available across multiple trials.

Comparisons from incremental or component-based designs were excluded where the trial arms progressively added behavioural components and no arm provided a stable control condition. In these cases, the resulting analysis estimates the contribution of individual components rather than the intervention–control effect targeted by the meta-analysis.

Omnibus tests

Trial results reported only through omnibus tests, such as an overall analysis of variance across several trial arms, were not included because these tests do not identify the direction or magnitude of individual treatment–control comparisons. Outcomes were included where sufficient information was available to recover an eligible treatment-control comparison and calculate an effect estimate.

These rules reflect the pre-specified treatment–control estimand and the outcome-level contrast-selection framework in the methodology plan.

Data extraction and preparation

Trial information was extracted from published BETA reports, technical appendices and supporting analysis materials. The extracted data captured the trial design, intervention and control conditions, outcomes, sample sizes, statistical results and characteristics used in the moderator analyses.

The source data were processed using reproducible R scripts. Data preparation included:

- standardising variable names and category labels
- creating identifiers for projects, trials and individual effect estimates
- converting relevant fields to numeric or date formats
- harmonising the coding of outcome types, expected effect directions and 1- or 2-sided hypothesis tests
- standardising binary outcomes reported as either percentages or proportions

- checking and correcting the ordering of confidence interval bounds
- identifying the type of variability estimate reported, such as, a group-level standard deviation or a regression standard error.

Where the source material did not include the treatment and control mean statistics but a control-group outcome and an average treatment effect was available, the treatment mean was derived from those values. These cases were identified as derived estimates in the dataset.

The preparation R script also created the treatment-minus-control difference used in subsequent effect-size calculations.

Effect-size calculation

Effect estimates were converted to Cohen's *d* measure of size effect so results measured in different units could be compared on a common scale. The calculation used the most direct calculation route available for each outcome, with treatment and control group results preferred over confidence intervals, and confidence intervals preferred over p-values.

Continuous outcomes

For continuous outcomes, effect sizes were initially calculated as Hedges' g using treatment and control group means, standard deviations and sample sizes. Hedges' *g* applies a small-sample correction to the standardised mean difference and is the default effect-size metric used in many meta-analytic procedures (Borenstein et al., 2009). Because the meta-analysis was conducted and reported on the Cohen's d scale, Hedges' *g* estimates were converted back to Cohen's *d* before analysis. Sampling variances were then calculated on the Cohen's *d* scale.

Reported regression standard errors were not treated as group-level standard deviations. Instead, where a reported standard error represented uncertainty around a group mean, it was converted to a standard deviation for a continuous outcome using the relevant group sample size.

Binary outcomes

For binary outcomes, treatment and control proportions were first standardised to a common scale from 0 to 1. Differences between proportions were transformed using Cohen's *h*, based on an arcsine square-root transformation, and then converted to the standardised mean difference scale. This placed binary and continuous outcomes on a common effect-size scale.

Selection of the effect-size calculation method

For each outcome, the analysis selected the most direct effect-size calculation available. Treatment and control group summary statistics, including means, standard deviations and sample sizes, were preferred where available. Where a direct calculation was not possible, effect sizes were derived from reported confidence intervals. Effect sizes could also be recovered from reported p-values as a final fallback approach, although this was not required for any effect estimate included in the final analysis. The calculation method used for each effect-size estimate was retained in the analysis dataset to support quality assurance and sensitivity checks.

Estimates derived from confidence intervals

Where the information required for a direct calculation was unavailable, effect sizes were derived from reported confidence intervals. Confidence intervals were converted to standard errors using the appropriate critical value, accounting for the available degrees of freedom and whether the interval related to a 1- or 2-sided test. The standard error and reported treatment–control difference was then used to derive a test statistic and Cohen's *d* effect size estimates.

The calculation framework also allowed effect sizes to be recovered from *p*-values as a final approach. However, this route was not required for any effect estimate included in the final analysis.

Direction of effects

All effects were initially calculated as the treatment-group outcome minus the control-group outcome. The sign was then aligned with the prespecified direction of the hypothesis:

- for outcomes expected to increase, a positive effect indicated an increase relative to control
- for outcomes expected to decrease, the sign was reversed so that a positive effect indicated a decrease relative to control
- where no direction was prespecified, the observed treatment-minus-control sign was retained.

Under this convention, positive values generally indicated movement in the hypothesised direction and negative values indicated movement in the opposite direction. Effects without a prespecified hypothesised direction were not classified as having worked or backfired and therefore retained their observed treatment-minus-control sign.

Sampling variances

To ensure an overall more robust meta-analysis, more precise estimates were given more weight by calculating a sampling variance for each effect estimate for use in inverse-variance weighting.

Extremely small estimated variances can give individual results implausibly large weights. A minimum variance threshold was therefore applied where necessary. The threshold was set as the larger of 1×10^{-6} or a lower-tail quantile of the observed variance distribution, ensuring that unusually small variance estimates could not dominate the inverse-variance weighting. Only one effect estimate required this adjustment.

Primary meta-analytic model

The primary analysis used a 3-level random-effects model to estimate the average effect while accounting for multiple effect estimates within trials and multiple trials within projects. A random-effects model was used because the effectiveness of behavioural interventions was expected to vary across trials and projects. The 3-level structure also accounted for the dependence among effect estimates from the same trial and allowed heterogeneity to be estimated at both the trial and project levels.

The model accounted for variation at 3 levels:

- **effect-estimate level** *(i)* - the statistical uncertainty around each result

- **trial level** *(j)* - differences between trials within the same BETA project
- **project level** *(k)* - differences between BETA projects.

The model can be expressed as:

$$d_{ijk} = \mu + u_k + v_{jk} + e_{ijk}$$

where:

d_{ijk} is effect estimate *i* from trial *j* in project *k*

μ is the estimated average effect across BETA's work

u_k is the random effect for project *k*

v_{jk} is the random effect for trial *j* within project *k*

e_{ijk} is sampling error, with variance fixed to the sampling variance of each effect estimate.

The model was estimated using restricted maximum likelihood. As noted above, effect estimates with smaller sampling variances received greater weight. The random-effects u_k allows the underlying effect to vary between projects, while v_{jk} allows the underlying effect to vary between trials within projects.

Robust statistical inference

Some trials contributed several related effect estimates because they measured more than one outcome or compared several intervention groups with the same control group. These estimates may be correlated because they are based on overlapping groups of participants.

The analysis therefore used trial-clustered robust standard errors. This reduced the risk of overstating the precision of the average effect when several related estimates came from the same trial. Sensitivity analyses also tested whether the result changed under different assumptions about the strength of these correlations (see Sensitivity section below for more details).

Heterogeneity

The model estimated how much effects differed between BETA projects and between trials within projects. A 95% prediction interval was also calculated. This shows the range of effects that could plausibly be observed in a comparable new trial, based on the variation observed across the existing evidence base.

Results

Primary meta-analysis

The primary 3-level random-effects model estimated a small positive mean effect across BETA's portfolio (d = 0.071). This effect corresponds approximately to an additional 10 to 20 people per 1,000 taking the intended action compared with what would have happened without the intervention, depending on the baseline rate of the behaviour.

However, substantial heterogeneity and a prediction interval spanning negative and positive effects indicate considerable variation across BETA's trials and projects (Table 1).

This result is in line with recent international syntheses of the impact of other behavioural policy interventions. For example, Kaiser et al. (2025) found that behaviour-change interventions produced a small positive average effect after adjustment for publication bias (d= 0.063), accompanied by substantial unexplained heterogeneity.

Table 1. Primary 3-level meta-analysis results – measure

Measure	Estimate
Effect estimates	148
Trials	52
Projects	40

Table 2. Primary 3-level meta-analysis results – mean effect

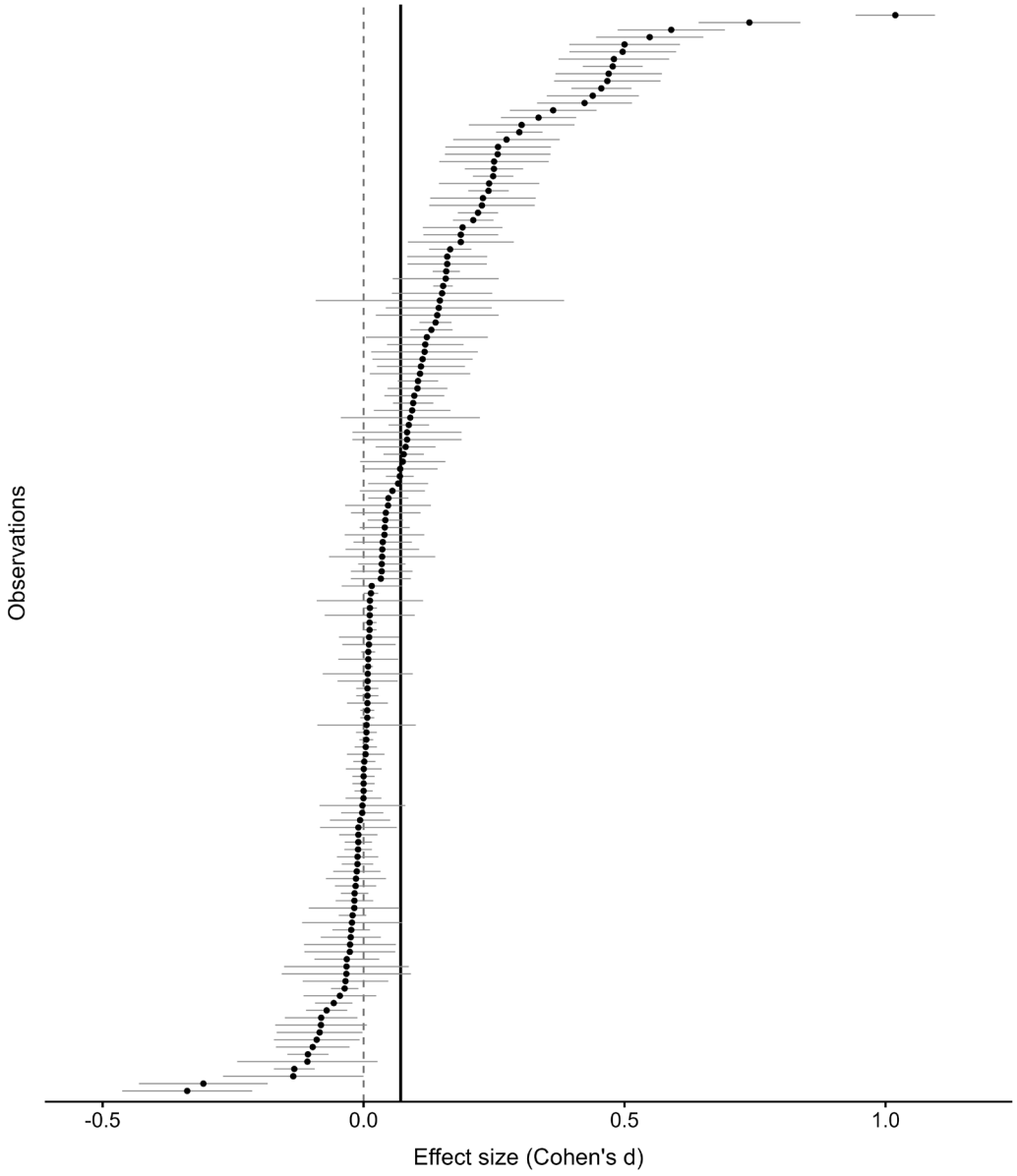
Mean effect	Estimate
Cohen's *d*	0.071
95% confidence interval	0.034–0.108
z-statistic	3.83
p-value	< .001

Table 3. Primary 3-level meta-analysis results – variation in effects

Variation in effects	Estimate
Total I^2	98.27%
Project-level I^2	32.6%
Trial-level I^2	65.20%
95% prediction interval	−0.191 to 0.333

Note: Positive effect sizes indicate movement in the prespecified hypothesised direction. The confidence interval uses trial-clustered robust standard errors. The prediction interval estimates the range of true effects that may plausibly occur in a comparable new trial, allowing for estimated heterogeneity.

Figure 2 Distribution of effect estimates included in the meta-analysis



Note: Each point represents an individual effect estimate included in the meta-analysis and the horizontal lines show 95% confidence intervals. The vertical black line shows the estimated average effect across BETA's portfolio (d = 0.07). The dashed line indicates no effect (that is, an effect size of zero).

Bayesian analysis

A Bayesian 3-level random-effects model was used as an additional check on the primary findings. This model analysed the same effect estimates and reflected the same 3-level structure of BETA's trials. Unlike the primary analysis, the Bayesian model allowed us to directly estimate the probability that the mean effect was positive or exceeded a specified threshold.

The Bayesian estimate closely matched the primary result (Table 2, Table 3). There was a greater than 99.9% probability that the mean effect across BETA's portfolio was positive and an 86% probability that it exceeded *d* = 0.05.

There was substantially greater uncertainty about the effect likely to occur in a comparable new context. The model estimated a 69% probability that the true effect would be positive and a 57% probability that it would exceed *d* = 0.05. The wide credible interval, which included both negative and positive effects, reflects the substantial variation observed across BETA's portfolio.

Table 4. Bayesian estimates of the mean effect and effect in a new context

Quantity	Estimate	95% credible interval
Mean effect, Cohen's d	0.072	0.031 to 0.113
Estimated true effect in a new context	0.073	−0.211 to 0.351

Table 5. Posterior probabilities

Quantity	Probability
Mean effect is positive	99.96%
Mean effect exceeds *d* = 0.05	86.01%
Estimated true effect in a new context is positive	69.44%
Estimated true effect in a new context exceeds *d* = 0.05	56.74%
Estimated true effect in a new context is below *d* = −0.05	19.38%

Note: The threshold of d = 0.05 was used as the smallest effect size of interest. In the Bayesian analysis, a credible interval describes the range containing 95% of the posterior distribution under the model and specified priors.

Prior sensitivity and model diagnostics

The estimated mean effect was similar under more sceptical and broader prior assumptions. The estimated true effects across in a new context, referred to as the posterior mean effects, ranged from *d* = 0.070 to *d* = 0.073, while the probability that the mean effect exceeded *d* = 0.05 ranged from 84% to 86%.

Standard diagnostic checks indicated that the Bayesian models converged successfully and produced a sufficient number of effective posterior samples. No divergent transitions or maximum tree-depth warnings were identified.

Moderator analyses

We also undertook moderator analyses to understand why estimated effect sizes differed across trials and projects. Moderator analyses examined the following characteristics:

- behavioural domain (finance, environment, prosocial behaviours that primarily benefit others or society, health and other policy domains, see Table 5)
- intervention category (the broad intervention approach: decision information, decision structure, or decision assistance, see Table 6)
- behavioural technique the specific behavioural technique implemented within those broader intervention categories, see Table 7
- experiment type (survey or field experiment, see Table 8) and

- trial year.

To prevent trials with many outcomes or intervention arms from having disproportionate influence in the moderator analyses, related estimates were first combined within trials using generalised least squares (GLS), assuming a within-trial correlation of ρ = 0.50. Random-effects meta-regressions were then estimated using restricted maximum likelihood.

None of the omnibus tests were statistically significant (Table 4). That is, the analysis did not identify clear evidence that the moderators examined systematically explained variation in estimate effects. Further, although the size of the mean effects differed across some subgroups of the moderating factors, the uncertainty around these estimates was substantial and several subgroups contained relatively few trials (Tables 5-8).

This is in line with recent second-order international syntheses of behavioural interventions which found substantial uncertainty remaining regarding their effectiveness across contexts (Hu et al., 2025; Kaiser et al., 2025).

Table 6. Omnibus tests of moderators

Moderator	Trials	Q_M	df	*p*-value
Behavioural domain	52	3.02	4	.555
Intervention category	52	2.15	2	.342
Behavioural technique	52	12.47	8	.132
Experiment type	51	2.54	1	.111
Trial year	50	0.00	1	.945
Intervention category × domain	52	4.40	9	.883
Behavioural technique × domain	52	23.53	25	.547

Note: QM tests whether effect sizes differ across moderator values. For categorical moderators, this corresponds to differences between levels of the moderator. Trial year was modelled as a continuous variable to assess whether effect sizes changed systematically over time rather than comparing individual years as separate categories. Moderator analyses were exploratory and no adjustment was made for multiple testing. This refers to the fact that several moderator analyses were examined, increasing the probability of observing a statistically significant result by chance alone. Df denotes the degrees of freedom associated with each test.

Behavioural domain

Table 7. Mean effects by behavioural domain

Domain	Trials	Cohen's d	95% confidence interval
Finance	19	0.109	0.034 to 0.184
Environment	12	0.053	−0.020 to 0.127
Prosocial	9	0.029	−0.060 to 0.118
Other	8	0.031	0.000 to 0.063
Health	4	0.041	−0.024 to 0.105

Intervention category

Table 8. Mean effects by intervention category

Category	Trials	Cohen's d	95% confidence interval
Decision information	34	0.054	0.017 to 0.092
Decision structure	11	0.059	0.003 to 0.115
Decision assistance	7	0.132	−0.055 to 0.318

Behavioural technique

Table 9. Mean effects by behavioural technique

Technique	Trials	Cohen's *d*	95% confidence interval
Make information visible	14	0.093	0.034 to 0.152
Translate information	12	0.041	−0.011 to 0.094
Change option-related effort	9	0.134	−0.011 to 0.280
Provide social reference	7	0.044	−0.009 to 0.097
Provide reminders	4	0.029	−0.074 to 0.132
Change range or composition	3	0.083	0.000 to 0.166
Change choice defaults	1	—	Not estimated
Change option consequences	1	—	Not estimated
Facilitate commitment	1	—	Not estimated

Experiment type

Table 10. Mean effects by experiment type

Experiment type	Trials	Cohen's *d*	95% confidence interval
Survey experiment	29	0.092	0.031 to 0.153
Field experiment	22	0.032	0.004 to 0.059

Note: Subgroup estimates are descriptive and should therefore be interpreted cautiously because the omnibus moderator tests in Table 4 did not detect statistically significant differences between subgroup levels. Technique-level estimates in Table 7 were not calculated for categories represented by fewer than 2 trials because there were insufficient data to estimate a subgroup effect.

Bayesian moderator analysis

Bayesian analyses supported the frequentist conclusion that none of the characteristics examined reliably explained variation in effects. The clearest directional pattern was for experiment type: there was a 93% probability that effects were larger in survey experiments than in field experiments, although the 95% credible interval for the difference included zero. There was little indication that the mean effect size varied by trial year.

Overall, the evidence remained too uncertain to identify reliable moderators, particularly for subgroups represented by relatively few trials.

Robustness testing

Sensitivity analyses

Sensitivity analyses tested whether the estimated mean effect changed under alternative approaches to handling related results from the same trial.

The main sensitivity analysis combined estimates within trials using generalised least squares (GLS), which accounts for their assumed correlation between effect-size estimates (ρ) and gives greater weight to more precise estimates. Additional sensitivity tests modelled one estimate per trial, and a 2-stage aggregation (that aggregated results first to the trial level and then to the project level) (Table 9).

The estimated mean effect remained positive across all modelling approaches, ranging from $d = 0.058$ to $d = 0.077$ (Table 9). All confidence intervals excluded zero and substantially overlapped with the confidence interval from the primary model. This indicates that the overall finding was not driven by trials contributing multiple effect estimates or by the particular hierarchical structure used in the primary analysis.

Table 11. Sensitivity analyses of the mean effect

Analysis	Cohen's *d*	95% confidence interval	Units included
Primary 3-level model	0.073	0.035 to 0.110	148 effects
GLS aggregation, ρ = 0.50	0.068	0.031 to 0.105	52 trials
One estimate per trial	0.077	0.040 to 0.115	52 trials
2-stage aggregation	0.058	0.024 to 0.091	40 projects

Note: The 2-stage analysis first combined effects within trials and then combined trials within projects. All models were estimated using restricted maximum likelihood.

Because the correlations between effect estimates from the same trial were not available, the GLS analysis was repeated using assumed correlations of ρ = 0.30, 0.50 and 0.70 (Table 10). The pooled mean effect size (Cohen's d) changed very little across the assumed correlations, indicating that the GLS result was not sensitive to the specific assumed level of correlation between effect-size estimates.

Table 12. GLS estimates across assumed within-trial correlations

Assumed correlation (ρ)	Cohen's d	95% confidence interval
0.30	0.070	0.033 to 0.107
0.50	0.068	0.031 to 0.105
0.70	0.066	0.028 to 0.103

Assessment of small-trial effects

We examined whether effect estimates were related to their precision using a funnel plot and a statistical test of small-trial effects methods.

Funnel-plot asymmetry can indicate underrepresentation of smaller trials with null or negative findings but can also arise from genuine differences between trials, particularly when

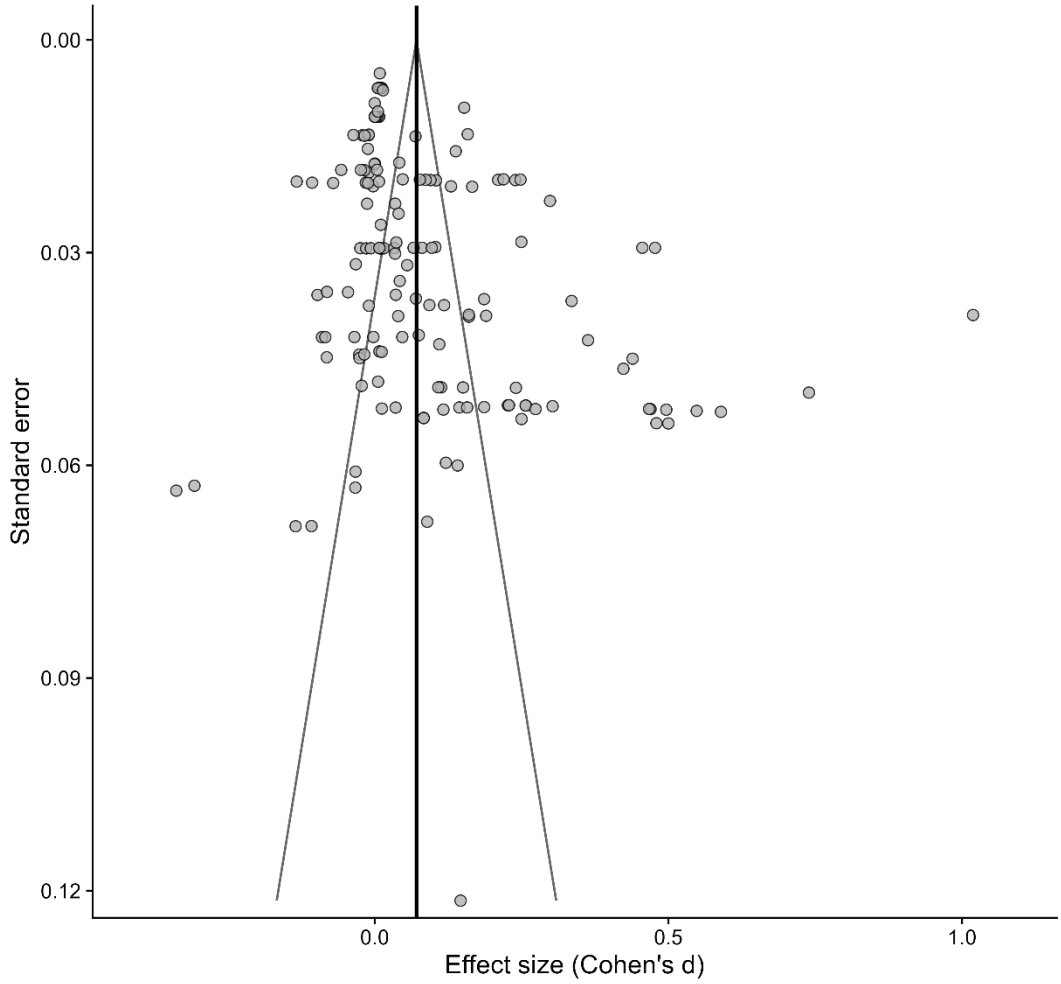
interventions, outcomes and settings are diverse. The funnel plot showed some asymmetry (Figure 3).

The Precision-Effect Test (PET) model found an association between effect size and standard error (b = 4.69, p = 0.021), while the corresponding Precision-Effect Estimate with Standard Error (PEESE) association was weaker (b = 52.31, p = 0.095).

Taken together, these results suggest that less precise effect estimates tended to have a larger effect size, although the evidence was not consistent across methods. This pattern is unlikely to reflect conventional publication bias because the analysis included BETA's completed trial portfolio rather than only trials published externally. It may instead reflect genuine differences between trials, substantial heterogeneity, or dependence among estimates from the same trial.

The funnel plot and PET–PEESE results should therefore be treated as diagnostic evidence of small-trial effects particularly because some trials in our meta-analysis contributed multiple related effect estimates, and not evidence that null or negative BETA trials were unpublished.

Figure 3 Funnel plot of effect estimates and standard errors



Note: Each point represents an effect estimate. The vertical line shows the estimated mean effect from the primary meta-analysis (d = 0.07). Funnel-plot asymmetry may reflect small-trial effects,

heterogeneity or dependence between estimates and should not be interpreted as evidence of publication bias on its own.

Limitations

The meta-analysis methodology was not preregistered. The analysis plan was refined during evidence extraction and methodological development, although the final approach was documented before the final models were run. This materially reduces the risk that decisions were made in response to the final results, but it does not provide the same protection against analytic flexibility as a fully preregistered protocol.

The evidence-base included outcomes reported in different formats, so effect sizes could not always be calculated in the same way. As outlined above, where group-level results were unavailable, estimates were recovered from reported confidence intervals, and binary and continuous outcomes were converted to a common standardised scale. These procedures required assumptions and may have introduced some additional measurement error, although the most direct calculation method available was used for each outcome.

Some trials contributed multiple effect estimates, including comparisons that reused a common control group. This created dependence between estimates that could not be modelled exactly as the underlying correlations were generally unavailable. The 3-level random effects model and trial-clustered robust standard errors reduced the risk of overstating precision, but they relied on assumptions about the dependence structure and may not have fully accounted for all the correlation between effect estimates. Sensitivity tests showed that the estimated mean effect remained positive and changed little across alternative modelling approaches, suggesting that the overall findings were not driven by the specific assumptions used to account for dependence.

Furthermore, the funnel plot and PET-PEESE analyses provided some evidence consistent with small-trial effects, although the findings were not consistent across methods and may also reflect substantial heterogeneity or dependence among effect estimates. These results should therefore be interpreted cautiously.

References

Borenstein M, Hedges LV, Higgins JPT and Rothstein HR (2009) *Introduction to meta-analysis*, Wiley.

Della Vigna S and Linos E (2022) *RCTs to scale: comprehensive evidence from two nudge units*, Econometrica, 90(1):81–116.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ and Welch VA (editors) (2022) *Cochrane handbook for systematic reviews of interventions*, version 6.3, Cochrane.

Hu, B., Xia, Z., Guo, Q., Lu, C., Constantino, S. M., and Ju, X. (2025) *Assessing nudge impact: A comprehensive second-order meta-analysis*, *Journal of Behavioral Decision Making, 38*(5), e70053. https://doi.org/10.1002/bdm.70053 [colab.ws], [hubo-home.com]

Kaiser T, Kloidt J, Mata J and Hertwig R (2025) *A meta-meta-analysis of behavior change interventions: Two tales of behavior change,* CESifo Working Paper No. 11863,. CESifo. https://www.ifo.de/DocDL/cesifo1_wp11863.pdf [ifo.de]

Mertens S, Herberz M, Hahnel UJJ and Brosch T (2022) *The effectiveness of nudging: a meta-analysis of choice architecture interventions across behavioral domains*, Proceedings of the National Academy of Sciences, 119(1):e2107346118.

Morris SB (2002) 'Combining effect sizes in meta-analysis with repeated measures and independent-groups designs', *Psychological Methods*, 7(1):105–125.

Van den Noortgate W, López-López JA, Marín-Martínez F and Sánchez-Meca J (2013) 'Meta-analysis of multiple outcomes: a multilevel approach', *Behavior Research Methods*, 45:576–594.

List and links to included trials

Project title	Link to report
Increasing participation in the AMEP	BETA project page
Supporting consumer decision-making about add-on insurance	BETA project page
Nudge vs Superbugs: Reducing the overprescribing of antibiotics	BETA project page
Testing incentives to encourage asbestos removal	BETA project page
Building Persistent Compliance with Labour Law	BETA project page
Prompting for a better deal: Understanding and overcoming barriers to financial product switching	BETA project page
Prompting for a better deal: Understanding and overcoming barriers to financial product switching	BETA project page
Helping people understand the outcomes of bankruptcy	BETA project page
Helping people understand the outcomes of bankruptcy	BETA project page
Improving energy bills	BETA project page
Improving energy bills	BETA project page
Improving government confirmation processes using SMS	BETA project page
Counting on us: How start dates affect Census participation	BETA project page
Testing communications strategies for the 2026 Census	BETA project page
Credit when it's due: Reducing credit card debt	BETA project page
password123: Applying behavioural insights to cyber security advice	BETA project page
Using behavioural insights to boost the impact of cyber security alerts	BETA project page
password123: Applying behavioural insights to cyber security advice	BETA project page
password123: Applying behavioural insights to cyber security advice	BETA project page
Encouraging survey participation to evaluate social service program outcomes	BETA project page
Making it less taxing - better compliance with deferred GST	BETA project page
Making it less taxing - better compliance with deferred GST	BETA project page

Increasing disability identification in the Australian Public Service	BETA project page
Empowered to switch: Energy made easier	BETA project page
Redesigning electricity bills to support consumer engagement	BETA project page
Simplifying energy fact sheets to improve consumer understanding	BETA project page
Energy labels that make cents: Testing the effect of appliance energy rating labels	BETA project page
Paid family and domestic violence leave	BETA project page
Strengthening Students' Resilience	BETA project page
The workplace experience of hospitality apprentices	BETA project page
Increasing transparency of online foreign exchange calculators	BETA project page
Applying behavioural insights to online wagering	BETA project page
Improving organ donor registration among young adults	BETA project page
Paid Parental Leave decisions	BETA project page
Introducing behavioural insights to the Personal Property Securities Register	BETA project page
Effective use of SMS to encourage timely reporting behaviour using digital channels	BETA project page
Supporting retirees in retirement income planning	BETA project page
Unconscious bias in Australian Public Service shortlisting processes	BETA project page
Increasing workplace giving	BETA project page
Improving tax compliance for income tax returns filed through tax agents	BETA project page
Women in Economics Student Mentoring Program	BETA project page
Informing the design of the YourSuper Comparison Tool	BETA project page
Home energy upgrades: empowering homeowners to make informed decisions	BETA project page

Tell us how you're using this report via a short survey below:

Other uses

Enquiries regarding this license and any other use of this document are welcome at:

Managing Director
Behavioural Economics Team of the Australian Government
Department of the Prime Minister and Cabinet
1 National Circuit
Barton ACT 2600
Email: beta@pmc.gov.au

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Department of the Prime Minister and Cabinet or the Australian Government.

Research team

Other (current and former) staff who contributed to the report were: Dr Angela Bender, Dr Bethany Jones, Michael Ridgway, Laura Bennetts-Kneebone and Dr Thomas Valentini.

Acknowledgements

Thank you to BETA's Academic Advisory Panel for their support and valuable contribution in making this project happen.

Use of Artificial Intelligence (AI)

During preparation of this report, the authors used Microsoft Copilot/ChatGPT to assist with drafting, editing, and code development. The authors reviewed, tested, and revised all AI-assisted content and analysis code and take full responsibility for the final report and any associated analyses, results, or software outputs.

ISBN 978-1-925365-95-5 *10 years of BETA: Meta-analysis of our behavioural trials*

Copyright Notice

With the exception of the Commonwealth Coat of Arms, this work is licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0) (https://creativecommons.org/licenses/by/4.0/)



Third party copyright

Wherever a third party holds copyright in this material, the copyright remains with that party. Their permission may be required to use the material. Please contact them directly.

Attribution

This publication should be attributed as follows:

© Commonwealth of Australia, Department of the Prime Minister and Cabinet, 1*0 years of BETA: Meta-analysis of our behavioural trials*, 2026.

Use of the Coat of Arms

The terms under which the Coat of Arms can be used are detailed on the following website: https://pmc.gov.au/cca

Australian Government

BETA

Behavioural Economics Team of the Australian Government

General enquiries beta@pmc.gov.au

Media enquiries media@pmc.gov.au

Find out more www.pmc.gov.au/BETA