Package 'magentabook'

September 8, 2026

Title HM Treasury Magenta Book Policy Evaluation Primitives

Version 0.1.1

Description Implements policy evaluation primitives from HM Treasury Magenta Book guidance (HM Treasury, 2026): theory of change and log-frame construction, evaluation planning and stakeholder mapping, power and minimum-detectable-effect calculations for randomised designs (including cluster and stepped-wedge designs following Hussey and Hughes (2007) <doi:10.1016/j.cct.2006.05.007> and Hemming et al. (2015) <doi:10.1136/bmj.h391>), Maryland Scientific Methods Scale ratings, structured confidence ratings, light-weight difference-in-differences and interrupted-time-series estimators (Bernal et al. (2017) <doi:10.1093/ije/dyw098>) with cluster-robust standard errors (Cameron and Miller (2015) <doi:10.3368/jhr.50.2.317>), pre-treatment balance checks (Stuart (2010) <doi:10.1214/09-STS313>), and cost-effectiveness analysis (cost per outcome, incremental cost-effectiveness ratio, acceptability curves, incremental net benefit, quality-adjusted and disability-adjusted life years). Designed as the evaluation companion to the appraisal package 'greenbook'. Bundled rubric and reference tables carry vintage metadata for reproducibility. Aligned with the May 2026 republication of the Magenta Book.

Depends R (>= 4.1.0)

License MIT + file LICENSE

Encoding UTF-8

Language en-US

RoxygenNote 7.3.3

Imports cli (>= 3.6.0), stats, utils

Suggests testthat (>= 3.0.0), knitr, rmarkdown, openxlsx, officer, flextable, pwr, sandwich, swCRTdesign, BCEA, cobalt

Config/testthat/edition 3

URL https://github.com/charlescoverdale/magentabook

BugReports https://github.com/charlescoverdale/magentabook/issues

VignetteBuilder knitr

NeedsCompilation no

Author Charles Coverdale [aut, cre]

Maintainer Charles Coverdale <charlesfcoverdale@gmail.com>

Repository CRAN

Date/Publication 2026-09-08 07:30:02 UTC

Contents

mb_assumptions . . . 3
mb_balance_table . . . 4
mb_cea . . . 5
mb_ceac . . . 6
mb_cluster_design . . . 7
mb_cmo . . . 8
mb_confidence . . . 9
mb_confidence_summary . . . 11
mb_contribution_claim . . . 12
mb_counterfactual . . . 13
mb_daly . . . 14
mb_data_versions . . . 15
mb_did_2x2 . . . 16
mb_evaluation_plan . . . 17
mb_evaluation_report . . . 19
mb_event_study . . . 20
mb_icc_reference . . . 22
mb_icer . . . 23
mb_inb . . . 24
mb_its . . . 25
mb_logframe . . . 26
mb_mde . . . 27
mb_power . . . 28
mb_qaly . . . 30
mb_questions . . . 31
mb_sample_size . . . 32
mb_schedule_table . . . 33
mb_sms_explain . . . 34
mb_sms_rate . . . 35
mb_stakeholders . . . 36
mb_stepped_wedge . . . 37
mb_theory_of_change . . . 39
mb_to_excel . . . 40
mb_to_latex . . . 41
mb_to_word . . . 42

Index **44**

mb_assumptions *Build a structured assumption register*

Description

Captures one or more assumptions from a theory of change in a tidy register, with the level they sit at, the supporting evidence (or its absence), and a criticality rating.

Usage

```
mb_assumptions(
  level,
  description,
  evidence = NA_character_,
  criticality = "medium"
)
```

Arguments

level	Character vector. The theory-of-change level the assumption sits at. One of "inputs", "activities", "outputs", "outcomes", "impact".
description	Character vector. Plain-English statement of the assumption.
evidence	Optional character vector. Source or rationale for believing the assumption holds. Defaults to NA.
criticality	Character vector. One of "low", "medium", "high". Failure of high-criticality assumptions invalidates the causal chain.

Value

An mb_assumption_register data frame with columns level, description, evidence, criticality.

See Also

mb_theory_of_change(), mb_logframe().

Other theory of change: mb_logframe(), mb_theory_of_change()

Examples

```
mb_assumptions(
  level       = c("activities", "outcomes"),
  description = c("Workshops are well-attended",
                  "Skills uplift translates into job entry"),
  evidence    = c("Pilot attendance 80%",
                  "Indirect: similar programmes show 0.3 SD effect"),
  criticality = c("medium", "high")
)
```

`mb_balance_table` *Pre-treatment balance table*

Description

Computes a Magenta Book-standard balance check for pre-treatment covariates: by-arm mean and standard deviation, standardised mean difference (SMD), and a two-sample test of equality. The SMD is the unitless effect size most evaluators report; rules of thumb flag `|SMD| > 0.10` as a meaningful imbalance and `|SMD| > 0.25` as a serious imbalance.

Usage

```
mb_balance_table(treated, ..., data = NULL, threshold = 0.1)
```

Arguments

`treated`	Logical or 0/1 numeric vector identifying the treated unit. `TRUE` / `1` means treated.
`...`	Numeric or factor covariates to balance check. Names become row labels. May be passed as a data frame via the `data` argument.
`data`	Optional data frame. If supplied, `...` is ignored and every column other than `treated` is checked. Pass the `treated` argument as a column reference (e.g. `data$treat`) or as the column name in the data frame.
`threshold`	Numeric scalar. Absolute SMD threshold above which a row is flagged as imbalanced. Default `0.10`.

Details

For a numeric or 0/1 covariate X with treated mean $\bar{X}_T$, control mean $\bar{X}_C$, treated SD s_T, and control SD s_C, the standardised mean difference is

$$\text{SMD} = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{(s_T^2 + s_C^2)/2}}.$$

This is the equal-weighted pooled-SD form recommended by Stuart (2010) and Austin (2009) for propensity-score balance diagnostics. It differs from Cohen's d, which uses the degrees-of-freedom-weighted pooled SD $\sqrt{(s_T^2(n_T - 1) + s_C^2(n_C - 1))/(n_T + n_C - 2)}$; the two agree when $n_T = n_C$. magentabook ships a cross-validation test against `cobalt::bal.tab` which uses the same averaged-SD form.

Rules of thumb (Cohen 1988; Stuart 2010):

- `|SMD| < 0.10`: well balanced
- `0.10 <= |SMD| < 0.25`: meaningful imbalance, consider covariate adjustment
- `|SMD| >= 0.25`: serious imbalance, matching or weighting recommended

Magenta Book impact evaluation guidance recommends a balance table for any quasi-experimental design and as a sense-check even for randomised designs.

Value

An mb_balance_table data frame with columns covariate, mean_treated, mean_control, sd_treated, sd_control, n_treated, n_control, smd, p_value, imbalanced. Numeric and binary covariates use the pooled-SD SMD and a Welch two-sample t-test. Factor covariates are decomposed into one row per non-reference level using the level-indicator and a chi-squared test on the original factor.

References

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. Statistical Science 25(1). doi:10.1214/09-STS313.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statistics in Medicine 28(25). doi:10.1002/sim.3697.

HM Treasury (2026). The Magenta Book: Annex A, Analytical methods for use within an evaluation. Section A2.2 on propensity score matching, where balance diagnostics are the canonical check on whether the matched comparison group is exchangeable with the treated group prior to estimation. https://www.gov.uk/government/publications/the-magenta-book.

See Also

mb_did_2x2(), mb_questions().

Other planning: mb_counterfactual(), mb_evaluation_plan(), mb_questions(), mb_stakeholders()

Examples

```
set.seed(20260427)
n <- 400
treated <- rep(c(0, 1), each = n / 2)
age     <- rnorm(n, mean = 45 + 2 * treated, sd = 10)
female  <- rbinom(n, 1, 0.5)
income  <- rnorm(n, mean = 30000 + 1500 * treated, sd = 8000)
mb_balance_table(treated = treated, age = age, female = female, income = income)
```

---

mb_cea *Cost per unit of outcome*

Description

Computes a simple cost-effectiveness ratio: total cost divided by total outcomes delivered. Use mb_icer() for two-option comparisons.

Usage

```
mb_cea(cost, effect, label = NULL)
```

Arguments

cost	Numeric scalar or vector. Total cost (or per-period costs that will be summed).
effect	Numeric scalar or vector. Total outcomes delivered (or per-period outcomes that will be summed).
label	Optional character scalar. Name of the option.

Value

An mb_cea object.

See Also

mb_icer(), mb_ceac(), mb_inb().

Other cost-effectiveness: mb_ceac(), mb_daly(), mb_icer(), mb_inb(), mb_qaly()

Examples

```
mb_cea(cost = 1e6, effect = 250, label = "Workshop programme")
```

---

mb_ceac *Cost-effectiveness acceptability curve*

Description

For a single A-vs-B comparison with sampled (delta_cost, delta_effect) draws (e.g. from a probabilistic sensitivity analysis), returns the probability that B is cost-effective at each willingness-to-pay (WTP) value in wtp_grid.

Usage

```
mb_ceac(delta_cost, delta_effect, wtp_grid)
```

Arguments

delta_cost	Numeric vector. Sampled incremental costs of B relative to A.
delta_effect	Numeric vector, same length as delta_cost. Sampled incremental effects.
wtp_grid	Numeric vector of willingness-to-pay values (cost per unit of effect) at which to evaluate the curve.

Details

At each WTP value lambda, B is cost-effective if the incremental net benefit lambda * delta_effect - delta_cost > 0. The CEAC is the proportion of draws for which this is true.

Value

An mb_ceac object: a data-frame-like list with columns wtp, prob_cost_effective, plus n_draws and vintage.

References

Fenwick, E., Claxton, K., Sculpher, M. (2001). Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Economics 10(8). doi:10.1002/hec.635.

See Also

mb_inb(), mb_icer().

Other cost-effectiveness: mb_cea(), mb_daly(), mb_icer(), mb_inb(), mb_qaly()

Examples

```
set.seed(4)
delta_cost   <- rnorm(1000, mean = 50000, sd = 10000)
delta_effect <- rnorm(1000, mean = 2,     sd = 0.5)
mb_ceac(delta_cost, delta_effect, wtp_grid = seq(0, 100000, by = 10000))
```

---

mb_cluster_design *Cluster-RCT design effect*

Description

Computes the design effect (DEFF) for a parallel cluster randomised trial: how much the variance of the treatment effect inflates relative to an individually-randomised design with the same total sample size, due to within-cluster correlation.

Usage

```
mb_cluster_design(individuals_per_cluster, icc, n_clusters = NULL)
```

Arguments

individuals_per_cluster	Numeric. Number of individuals sampled per cluster (m).
icc	Numeric in [0, 1]. Intra-class correlation coefficient. Use mb_icc_reference() for plausible values.
n_clusters	Optional numeric. Number of clusters per arm. If supplied, returns effective sample size per arm in addition to the design effect.

Details

$$\text{DEFF} = 1 + (m - 1)\,\rho$$

where `m` is the cluster size and `rho` is the ICC. The effective sample size for power is `n_total / DEFF`.

Standard reference values for `rho` across UK policy domains are bundled and accessible via `mb_icc_reference()`.

Value

A list with elements `deff` and (if `n_clusters` supplied) `n_total_per_arm` and `n_effective_per_arm`.

References

Donner, A., Klar, N. (2000). Design and Analysis of Cluster Randomization Trials in Health Research. Arnold.

Hedges, L. V., Hedberg, E. C. (2007). Intraclass Correlation Values for Planning Group-Randomized Trials in Education. Educational Evaluation and Policy Analysis 29(1). doi:10.3102/0162373707299706.

See Also

`mb_icc_reference()`, `mb_stepped_wedge()`, `mb_sample_size()`.

Other power: `mb_icc_reference()`, `mb_mde()`, `mb_power()`, `mb_sample_size()`, `mb_stepped_wedge()`

Examples

```
mb_cluster_design(individuals_per_cluster = 30, icc = 0.05)
mb_cluster_design(individuals_per_cluster = 30, icc = 0.05, n_clusters = 20)
```

---

mb_cmo *Context-mechanism-outcome (CMO) configuration*

Description

Records one or more CMO configurations from a realist evaluation: the *contexts* in which a *mechanism* fires to produce an *outcome*, with optional supporting evidence.

Usage

```
mb_cmo(context, mechanism, outcome, evidence = NA_character_)
```

Arguments

`context`	Character vector. The contextual conditions needed for the mechanism to fire.
`mechanism`	Character vector. The underlying generative mechanism (typically a change in reasoning or resources).
`outcome`	Character vector. The observed outcome pattern.
`evidence`	Character vector. Citation, quote, or other evidence supporting the configuration. Default `NA`.

Details

Realist evaluation, developed by Pawson and Tilley (1997), seeks to answer "what works for whom in what circumstances and why" by surfacing CMO configurations rather than estimating average treatment effects. The Magenta Book lists realist evaluation as the principal theory-based approach for context-dependent interventions.

Value

An `mb_cmo` data frame with columns `context`, `mechanism`, `outcome`, `evidence`.

References

Pawson, R., Tilley, N. (1997). Realistic Evaluation. SAGE.

HM Treasury (2026). The Magenta Book: Annex A, Analytical methods for use within an evaluation. Section A1.2 on realist evaluation. https://www.gov.uk/government/publications/the-magenta-book.

See Also

`mb_contribution_claim()`.

Other realist: `mb_contribution_claim()`

Examples

```
mb_cmo(
  context   = c("High trust GP-patient relationships",
                "Low trust GP-patient relationships"),
  mechanism = c("Patients accept advice", "Patients ignore advice"),
  outcome   = c("Improved adherence", "No change in adherence"),
  evidence  = c("Smith et al. 2024 cohort study", "Smith et al. 2024")
)
```

---

mb_confidence *Structured Magenta Book confidence rating*

Description

Records a single confidence rating against the bundled rubric: high / medium / low, with explicit assessments of evidence strength, methodological quality, and generalisability, and a free-text rationale.

Usage

```
mb_confidence(
  rating = c("high", "medium", "low"),
  question,
  evidence_strength,
  methodological_quality,
  generalisability,
  rationale
)
```

Arguments

`rating`	Character scalar. One of `"high"`, `"medium"`, `"low"`.
`question`	Character scalar. The evaluation question this rating refers to.
`evidence_strength`	Character scalar. Plain-English description of the volume and quality of underlying studies.
`methodological_quality`	Character scalar. Plain-English description of design rigour and identifying assumptions.
`generalisability`	Character scalar. Plain-English description of how widely the findings travel across settings.
`rationale`	Character scalar. Free-text justification for the chosen rating.

Details

Magenta Book confidence ratings translate evidence into decision-grade summaries for ministers and senior officials. The bundled rubric (see `mb_schedule_table()` with table `"confidence"`) is *not a direct quotation from the Magenta Book*. It is a magentabook synthesis of cross-What-Works-Centre confidence-rating traditions: Education Endowment Foundation (5 padlocks), Early Intervention Foundation (Foundation Standards), College of Policing (1-5 scale), and the Justice Data Lab (red / amber / green). The three-level high / medium / low structure is designed for HMG decision-grade reporting and aligns with the value-for-money framing of the Magenta Book (HM Treasury, 2026, Chapter 3.6 and Annex A Section A3).

Value

An `mb_confidence` object: a list with the supplied fields plus the bundled-rubric `description` for the chosen rating, and `vintage`.

References

HM Treasury (2026). The Magenta Book: Central Government Guidance on Evaluation. Chapter 3.6 on value for money evaluation methods and Annex A Section A3. https://www.gov.uk/government/publications/the-magenta-book.

Education Endowment Foundation. Padlock evidence ratings.

Early Intervention Foundation (2021). Foundation Standards of Evidence.

See Also

mb_confidence_summary(), mb_sms_rate().

Other confidence: mb_confidence_summary()

Examples

```
mb_confidence(
  rating                = "medium",
  question              = "Did the policy raise employment",
  evidence_strength     = "One Level 4 DiD; one Level 3 matched cohort",
  methodological_quality = "Adequate; parallel trends plausible but limited pre-period",
  generalisability      = "Findings established in a single region",
 rationale         = "Effect direction consistent across two studies but limited replication"
)
```

---

mb_confidence_summary *One-page confidence summary across multiple ratings*

Description

Aggregates several mb_confidence ratings into a single summary object with a confidence count and the underlying ratings as a data frame.

Usage

```
mb_confidence_summary(...)
```

Arguments

...	One or more mb_confidence objects, or a single list of them.

Value

An mb_confidence_summary object: a list with n (total ratings), counts (named integer vector by rating), ratings (data frame), and vintage.

See Also

mb_confidence().

Other confidence: mb_confidence()

Examples

```
c1 <- mb_confidence(
  "high",   "Did employment rise",
  "Two Level 5 RCTs", "Strong; randomisation worked",
  "Tested in two regions", "Two RCTs both positive"
)
c2 <- mb_confidence(
  "medium", "Did wages rise",
  "One Level 4 DiD",  "Adequate; parallel trends plausible",
  "Single region",    "DiD effect positive but no replication"
)
mb_confidence_summary(c1, c2)
```

---

mb_contribution_claim *Contribution-analysis claim*

Description

Records a contribution claim with supporting and refuting evidence and an overall strength rating. Used in contribution-analysis-style theory-based evaluation, where causal inference comes from triangulating multiple evidence streams against a contribution story rather than from a counterfactual.

Usage

```
mb_contribution_claim(
  claim,
  evidence_for,
  evidence_against = character(0),
  strength = c("weak", "moderate", "strong")
)
```

Arguments

`claim`	Character scalar. The contribution claim being tested.
`evidence_for`	Character vector. Evidence supporting the claim.
`evidence_against`	Character vector. Evidence against the claim. Default `character(0)`.
`strength`	Character scalar. One of `"weak"`, `"moderate"`, `"strong"`. Reflects the analyst's overall judgement after weighing evidence for and against.

Value

An `mb_contribution_claim` object.

References

Mayne, J. (2008). Contribution Analysis: An approach to exploring cause and effect. ILAC Brief No. 16.

HM Treasury (2026). The Magenta Book: Annex A, Analytical methods for use within an evaluation. Section A1.4 on contribution analysis. https://www.gov.uk/government/publications/the-magenta-book.

See Also

mb_cmo().

Other realist: mb_cmo()

Examples

```
mb_contribution_claim(
  claim            = "The training programme contributed to higher employment",
  evidence_for     = c("Pre-post outcomes improved",
                       "Theory of change pathways visible in interviews"),
  evidence_against = "Macro labour market also improved",
  strength         = "moderate"
)
```

---

mb_counterfactual *Define a counterfactual*

Description

Records the comparison condition against which the policy effect is to be measured. The Magenta Book stresses that no impact evaluation is possible without an explicit counterfactual.

Usage

```
mb_counterfactual(
  definition,
 source = c("rct", "quasi-experimental", "theory-based", "comparator", "historical"),
  credibility = NA_character_
)
```

Arguments

definition	Character scalar describing the counterfactual: what would have happened in the absence of the policy.
source	Character scalar. Mechanism by which the counterfactual is constructed. One of "rct", "quasi-experimental", "theory-based", "comparator", "historical".
credibility	Character scalar. Plain-English assessment of how credible the counterfactual is.

Value

An mb_counterfactual object.

References

HM Treasury (2026). The Magenta Book: Annex A, Analytical methods for use within an evaluation. Section A2 on experimental and quasi-experimental methods (the counterfactual is the comparison group, time period, or unit that proxies what would have happened in the absence of the intervention). https://www.gov.uk/government/publications/the-magenta-book.

See Also

mb_evaluation_plan().

Other planning: mb_balance_table(), mb_evaluation_plan(), mb_questions(), mb_stakeholders()

Examples

```
mb_counterfactual(
  definition  = "Eligible non-applicants in the same year",
  source      = "quasi-experimental",
  credibility = "Moderate; selection on observables only"
)
```

---

mb_daly *Disability-adjusted life years (DALYs) accumulator*

Description

Sums years lived with disability (YLD) and years of life lost (YLL) across persons. DALY is the global-health analogue of QALY: lower is better.

Usage

```
mb_daly(yld, yll, persons = 1)
```

Arguments

yld	Numeric scalar or vector. Years lived with disability per person.
yll	Numeric scalar or vector. Years of life lost per person (e.g. life expectancy minus age at death).
persons	Numeric scalar. Number of persons. Default 1.

Details

$$\text{DALY} = \text{persons} \cdot \sum (YLD + YLL)$$

This implementation follows the Global Burden of Disease definition. Age-weighting and discounting are not applied by default (the IHME GBD removed both in the 2010 update); add a discount factor manually if your guidance still requires it.

Value

Numeric scalar. Total DALYs (YLD + YLL summed across persons).

References

Murray, C. J. L., Lopez, A. D. (1996). The Global Burden of Disease. Harvard University Press.

GBD 2019 Diseases and Injuries Collaborators (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019. The Lancet 396. doi:10.1016/S0140-6736(20)30925-9.

See Also

`mb_qaly()`.

Other cost-effectiveness: `mb_cea()`, `mb_ceac()`, `mb_icer()`, `mb_inb()`, `mb_qaly()`

Examples

```
mb_daly(yld = 2.5, yll = 8.0, persons = 100)
```

---

mb_data_versions *Vintage of bundled rubric and reference tables*

Description

Returns a data frame describing the source and last-updated date of every CSV bundled in `inst/extdata/`. Critical for reproducibility: every evaluation report can record the vintage of the rubrics and reference values used.

Usage

```
mb_data_versions()
```

Value

A data frame with columns `dataset`, `source`, `last_updated`, `notes`.

See Also

mb_schedule_table().

Other lookups: mb_schedule_table()

Examples

```
mb_data_versions()
```

---

mb_did_2x2 *Canonical 2x2 difference-in-differences estimator*

Description

Returns the simple two-period, two-group DiD estimate of an average treatment effect on the treated, with optional cluster-robust standard errors.

Usage

```
mb_did_2x2(y, treated, post, cluster = NULL, alpha = 0.05, quiet = FALSE)
```

Arguments

y	Numeric vector of outcomes.
treated	Logical or 0/1 numeric vector. TRUE / 1 if the unit is in the treated group, regardless of period.
post	Logical or 0/1 numeric vector. TRUE / 1 if the observation is in the post-treatment period.
cluster	Optional vector identifying clusters for cluster-robust standard errors (CR1 with finite-sample correction). If NULL, conventional OLS SEs are returned.
alpha	Numeric in (0, 1). Significance level for the confidence interval. Default 0.05.
quiet	Logical. If FALSE (default), the print method appends a one-line reminder that this is a canonical 2x2 DiD and points to specialist tooling for staggered adoption or heterogeneous treatment effects. Set to TRUE to suppress.

Details

Computes

$$\hat{\tau} = (\bar{Y}_{T,1} - \bar{Y}_{T,0}) - (\bar{Y}_{C,1} - \bar{Y}_{C,0})$$

which equals the coefficient on the treated:post interaction in $Y = \beta_0 + \beta_1 T + \beta_2 P + \tau(T \times P) + \epsilon$.

Cluster-robust SEs use the CR1 sandwich estimator with finite-sample correction $(G/(G-1)) \cdot (N-1)/(N-K)$, where G is the number of clusters, N the number of observations, and K the number of regressors (4).

For staggered adoption, heterogeneous treatment effects, or production estimation, use **fixest**, **did**, or **Synth**. This function is for the canonical 2x2 case only.

Value

An mb_did object: a list with estimate, se, t_stat, p_value, ci_low, ci_high, group means, cluster_robust, n, quiet, and vintage.

References

Card, D., Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. American Economic Review 84(4). doi:10.1257/aer.84.4.772.

Cameron, A. C., Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. Journal of Human Resources 50(2). doi:10.3368/jhr.50.2.317.

HM Treasury (2026). The Magenta Book: Annex A, Analytical methods for use within an evaluation. Section A2.7 on difference-in-difference. https://www.gov.uk/government/publications/the-magenta-book.

See Also

mb_its(), mb_event_study().

Other estimators: mb_event_study(), mb_its()

Examples

```
set.seed(1)
n <- 400
treated <- rep(c(0, 1), each = n / 2)
post    <- rep(c(0, 1), times = n / 2)
y       <- 0.5 * treated + 0.2 * post + 0.4 * treated * post + rnorm(n)
mb_did_2x2(y, treated, post)
```

---

mb_evaluation_plan *Aggregate evaluation plan*

Description

Composes the evaluation scope, questions, methods, timing, governance, and (optionally) budget into a single object suitable for review and export.

Usage

```
mb_evaluation_plan(
  scope,
  questions,
  methods,
  timing,
  governance,
  budget = NULL
)
```

Arguments

scope	Character scalar describing what the evaluation does and does not cover.
questions	An mb_questions object.
methods	Character vector of methods chosen for each type of question (e.g. c(impact = "RCT", process = "Realist interviews")). Names are used in the print method.
timing	Character vector or list describing the evaluation timeline (baseline, midline, endline, follow-up).
governance	Character vector or list describing oversight: steering group composition, peer review, data access.
budget	Optional numeric scalar (GBP) for total evaluation cost.

Value

An mb_plan object.

References

HM Treasury (2026). The Magenta Book: Central Government Guidance on Evaluation. Chapter 2 on evaluation scoping and Chapter 5 on managing an evaluation. https://www.gov.uk/government/publications/the-magenta-book.

See Also

mb_questions(), mb_counterfactual(), mb_stakeholders(), mb_evaluation_report().

Other planning: mb_balance_table(), mb_counterfactual(), mb_questions(), mb_stakeholders()

Examples

```
qs <- mb_questions(
  text = c("Did employment rise", "Was the policy implemented faithfully"),
  type = c("impact", "process")
)
mb_evaluation_plan(
  scope      = "GBP 50m skills programme, 2026-2029",
  questions  = qs,
  methods    = c(impact = "RCT", process = "Mixed methods"),
  timing     = c(baseline = "2026-Q1", endline = "2029-Q2"),
  governance = "Joint HMT / DfE steering group; peer review by What Works"
)
```

`mb_evaluation_report` *Aggregate evaluation report*

Description

Composes the components produced by other `magentabook` functions into a single report object: theory of change, evaluation plan, SMS ratings, confidence ratings, cost-effectiveness analyses. Any component may be omitted.

Usage

```
mb_evaluation_report(
  plan = NULL,
  toc = NULL,
  sms = NULL,
  confidence = NULL,
  cea = NULL,
  name = NULL
)
```

Arguments

`plan`	Optional `mb_plan` from `mb_evaluation_plan()`.
`toc`	Optional `mb_toc` from `mb_theory_of_change()`.
`sms`	Optional `mb_sms_rating` or list of them.
`confidence`	Optional `mb_confidence`, `mb_confidence_summary`, or list of `mb_confidence`.
`cea`	Optional `mb_cea`, `mb_icer`, or list of them.
`name`	Optional character scalar naming the evaluation.

Value

An `mb_report` object.

See Also

`mb_to_word()`, `mb_to_excel()`, `mb_to_latex()`.

Other reporting: `mb_to_excel()`, `mb_to_latex()`, `mb_to_word()`

Examples

```
toc <- mb_theory_of_change(
  inputs = "Funding", activities = "Workshops",
  outputs = "Attendees", outcomes = "Skills",
  impact = "Employment"
)
mb_evaluation_report(toc = toc, name = "Skills uplift evaluation")
```

`mb_event_study` *Simple event-study coefficients*

Description

Estimates a panel event-study with unit and time fixed effects and event-time dummies. Treatment time is fixed across treated units (no staggered adoption). Returns coefficients for `leads` periods before and `lags` periods after treatment, with the period immediately before treatment (`event_time = -1`) omitted as the reference category.

Usage

```
mb_event_study(
  y,
  unit,
  time,
  treatment_time,
  treated,
  leads = 3L,
  lags = 3L,
  cluster = NULL,
  quiet = FALSE
)
```

Arguments

`y`	Numeric vector of outcomes.
`unit`	Vector identifying units (panel ID).
`time`	Numeric vector of time indices.
`treatment_time`	Numeric scalar. The first treated period. Units with `treated = 0` (never-treated) are pure controls.
`treated`	Logical or 0/1 numeric vector indicating whether each observation belongs to a treated unit. The design requires at least some never-treated control units; without them the event-time dummies are collinear with the time fixed effects.
`leads`	Integer >= 0. Number of pre-treatment periods to include. Default `3`.
`lags`	Integer >= 0. Number of post-treatment periods. Default `3`.
`cluster`	Optional vector identifying clusters for cluster-robust standard errors (CR1 with finite-sample correction `(G/(G-1)) * (N-1)/(N-K)`). Common choice: pass `unit` to cluster at the unit level. If `NULL` (default), conventional OLS SEs are returned.
`quiet`	Logical. If `FALSE` (default), the print method appends a one-line reminder that this is a fixed-treatment- time event study and points to **fixest** (`sunab()`) or **did** for staggered adoption. Set to `TRUE` to suppress.

Details

Implements the canonical two-way fixed-effects event study:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{k \neq -1} \beta_k \mathbf{1}\{t - t^* = k, D_i = 1\} + \epsilon_{it}$$

For staggered adoption (units treated at different times), this specification is biased under treatment-effect heterogeneity. Use the heterogeneity-robust estimators of Callaway & Sant'Anna (2021) or de Chaisemartin & D'Haultfoeuille (2020), available in the **did**, **didimputation**, or **fixest** packages (`fixest::feols` with `sunab()`).

Standard errors are conventional OLS; for clustered inference use **sandwich** or **fixest**.

Value

An `mb_event_study` object: a list with `event_time`, `estimate`, `se`, plus `n`, `n_units`, `n_periods`, `treatment_time`, and `vintage`.

References

Callaway, B., Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. Journal of Econometrics 225(2). doi:10.1016/j.jeconom.2020.12.001.

de Chaisemartin, C., D'Haultfoeuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. American Economic Review 110(9). doi:10.1257/aer.20181169.

HM Treasury (2026). The Magenta Book: Annex A, Analytical methods for use within an evaluation. Section A2.7 on difference-in-difference (the event study is a time-resolved generalisation of two-period difference-in-difference). https://www.gov.uk/government/publications/the-magenta-book.

See Also

`mb_did_2x2()`, `mb_its()`.

Other estimators: `mb_did_2x2()`, `mb_its()`

Examples

```
set.seed(3)
n_units <- 50; n_periods <- 10; treat_time <- 6
panel <- expand.grid(unit = 1:n_units, time = 1:n_periods)
panel$treated <- as.integer(panel$unit <= 25)
panel$post    <- as.integer(panel$time >= treat_time)
panel$y <- 0.1 * panel$time + 0.5 * (panel$treated * panel$post) +
           rnorm(nrow(panel))
mb_event_study(
  y = panel$y, unit = panel$unit, time = panel$time,
  treatment_time = treat_time, treated = panel$treated,
  leads = 3, lags = 3
)
```

`mb_icc_reference` *Reference intra-class correlation values*

Description

Returns bundled reference ICC values for common UK policy domains and units of clustering. Use these for evaluation planning when domain-specific baseline data are not available.

Usage

```
mb_icc_reference(domain = NULL)
```

Arguments

`domain`	Optional character scalar. One of `"education"`, `"health"`, `"employment"`, `"local_government"`, `"criminal_justice"`, `"housing"`. If `NULL` (default), returns the entire reference table.

Details

Values are *reference* ICCs for planning purposes only. Wherever feasible, evaluators should compute domain-specific ICCs from baseline data before finalising sample size calculations.

Each row carries a `value_source` flag:

- `"table_quote"`: direct extraction of a specific row or value from a published table (cited table number in the `source` field).
- `"central_estimate"`: researcher synthesis of a plausible central value within the published range, used as a practitioner default in the absence of domain-specific baseline data.

At v0.1.0 every bundled row is `central_estimate`. Future versions will upgrade individual rows to `table_quote` as exact table-level citations are added. Treat the bundled values as a planning prior; verify against your own baseline ICC before relying on them in a published power calculation.

Value

A data frame with columns `domain`, `outcome`, `unit_of_clustering`, `icc_low`, `icc_central`, `icc_high`, `value_source`, `source`, `notes`.

References

Hedges, L. V., Hedberg, E. C. (2007). Educational Evaluation and Policy Analysis 29(1). doi:10.3102/0162373707299706.

Adams, G., Gulliford, M. C., Ukoumunne, O. C., Eldridge, S., Chinn, S., Campbell, M. J. (2004). Patterns of intra-cluster correlation from primary care research. Statistics in Medicine 23. doi:10.1002/sim.1764.

Campbell, M. K., Mollison, J., Grimshaw, J. M. (2000). Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. BMJ 321. doi:10.1136/bmj.321.7263.778.

See Also

mb_cluster_design(), mb_stepped_wedge().

Other power: mb_cluster_design(), mb_mde(), mb_power(), mb_sample_size(), mb_stepped_wedge()

Examples

```
mb_icc_reference()
mb_icc_reference("education")
```

---

mb_icer *Incremental cost-effectiveness ratio with dominance handling*

Description

Computes the ICER comparing option B to option A, with explicit handling of the four dominance regions:

- **A dominates** B (B costs more, delivers less): no ICER.
- **B dominates** A (B costs less, delivers more): no ICER; B is the obvious choice.
- **B more costly, more effective**: standard ICER positive.
- **B less costly, less effective**: ICER negative — B saves money at the expense of effect.

Usage

```
mb_icer(cost_a, effect_a, cost_b, effect_b, label_a = "A", label_b = "B")
```

Arguments

cost_a, effect_a
: Numeric scalars. Cost and effect of option A.

cost_b, effect_b
: Numeric scalars. Cost and effect of option B.

label_a, label_b
: Character scalars. Labels for the two options.

Details

The ICER is the cost per additional unit of outcome from switching from A to B:

$$\text{ICER} = (C_B - C_A)/(E_B - E_A)$$

If delta_effect is zero, the ICER is reported as Inf (when costs differ) or NaN (when costs are equal).

Value

An mb_icer object: a list with delta_cost, delta_effect, icer, dominance (one of "a_dominates", "b_dominates", "b_more_costly_more_effective", "b_less_costly_less_effective"), and labels.

References

HM Treasury (2026). The Magenta Book: Annex A, Analytical methods for use within an evaluation. Section A3.3 on cost-effectiveness analysis and Section A3.4 on cost utility analysis. https://www.gov.uk/government/publications/the-magenta-book.

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., Torrance, G. W. (2015). Methods for the Economic Evaluation of Health Care Programmes (4th ed.). Oxford University Press.

See Also

mb_cea(), mb_ceac(), mb_inb().

Other cost-effectiveness: mb_cea(), mb_ceac(), mb_daly(), mb_inb(), mb_qaly()

Examples

```
mb_icer(cost_a = 1e6, effect_a = 200, cost_b = 1.5e6, effect_b = 300,
        label_a = "Status quo", label_b = "Enhanced")
```

---

mb_inb *Incremental net benefit*

Description

Computes the incremental net benefit (INB) of B over A at a single willingness-to-pay threshold. Equivalent to the ICER framing on a monetary scale.

Usage

```
mb_inb(delta_cost, delta_effect, wtp)
```

Arguments

delta_cost	Numeric scalar. Incremental cost of B over A.
delta_effect	Numeric scalar. Incremental effect of B over A.
wtp	Numeric scalar. Willingness-to-pay per unit of effect (e.g. the NICE QALY threshold in a health context).

Details

$$\text{INB} = \lambda \cdot \Delta E - \Delta C$$

Equivalent to ICER comparison: INB > 0 iff ICER < WTP (when effect change is positive).

Value

Numeric scalar. INB in the units of `delta_cost`. INB > 0 means B is cost-effective at the supplied WTP.

See Also

`mb_ceac()`, `mb_icer()`.

Other cost-effectiveness: `mb_cea()`, `mb_ceac()`, `mb_daly()`, `mb_icer()`, `mb_qaly()`

Examples

```
mb_inb(delta_cost = 50000, delta_effect = 2, wtp = 30000)
```

---

mb_its *Interrupted time series via segmented regression*

Description

Fits a single-group interrupted time series model:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 P_t + \beta_3 (t - t^*) P_t + \epsilon_t$$

where `P_t` is 1 for `t >= t*` and `t*` is the intervention time. `beta_2` is the immediate level change at the intervention; `beta_3` is the change in slope.

Usage

```
mb_its(y, time, intervention_time, lag = 0L, quiet = FALSE)
```

Arguments

`y`	Numeric vector of outcomes ordered by `time`.
`time`	Numeric vector of time indices, same length as `y`.
`intervention_time`	Numeric scalar. The first time point considered post-intervention.
`lag`	Integer >= 0. Number of pre-intervention observations to drop near the intervention (transition period). Default `0`.
`quiet`	Logical. If `FALSE` (default), the print method appends a one-line reminder that this is a single-group segmented regression and points to specialist tooling for autocorrelation correction. Set to `TRUE` to suppress.

Details

Segmented regression assumes residuals are independent. For autocorrelated series, fit a Newey-West, Prais-Winsten, or ARIMA-error specification using **sandwich**, **nlme**, or **forecast**. This function is the canonical baseline.

Value

An mb_its object: a list with coefficients (named numeric), se (named numeric), level_change, slope_change, intervention_time, n, n_pre, n_post, and vintage.

References

Bernal, J. L., Cummins, S., Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: a tutorial. International Journal of Epidemiology 46(1). doi:10.1093/ije/dyw098.

Wagner, A. K., Soumerai, S. B., Zhang, F., Ross-Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. Journal of Clinical Pharmacy and Therapeutics 27. doi:10.1046/j.1365-2710.2002.00430.x.

HM Treasury (2026). The Magenta Book: Annex A, Analytical methods for use within an evaluation. Section A2.4 on interrupted time series analysis. https://www.gov.uk/government/publications/the-magenta-book.

See Also

mb_did_2x2(), mb_event_study().

Other estimators: mb_did_2x2(), mb_event_study()

Examples

```
set.seed(2)
time <- 1:48
y    <- 10 + 0.05 * time + ifelse(time >= 25, 2 + 0.1 * (time - 25), 0) + rnorm(48, sd = 0.5)
mb_its(y, time, intervention_time = 25)
```

---

mb_logframe *Convert a theory of change into a logframe*

Description

Pivots an mb_toc into a logframe table: one row per level, with optional indicators, means of verification, and risks columns. The May 2026 republication of the Magenta Book uses the term "logic model" for the underlying flow (inputs through to impact), but the tabular logframe layout (originating in DFID / FCDO and EU project management practice) remains widely used across UK evaluation reports and is the form produced here.

Usage

```
mb_logframe(toc, indicators = NULL, mov = NULL, risks = NULL)
```

Arguments

toc	An mb_toc object from mb_theory_of_change().
indicators	Optional named list. Names must be one of "inputs", "activities", "outputs", "outcomes", or "impact". Each element is a character vector of indicators.
mov	Optional named list, same convention. Means of verification per level (data source, survey, administrative record).
risks	Optional named list, same convention. Risks per level.

Value

An mb_logframe object: a data frame with columns level, description, and (if supplied) indicator, mov, risk. Multiple items per level are concatenated with "; ".

See Also

mb_theory_of_change().

Other theory of change: mb_assumptions(), mb_theory_of_change()

Examples

```
toc <- mb_theory_of_change(
  inputs = "Funding", activities = "Workshops",
  outputs = "Attendees", outcomes = "Skills",
  impact = "Employment"
)
mb_logframe(
  toc,
  indicators = list(outputs = "n attendees", outcomes = "skills score"),
  mov        = list(outputs = "attendance log", outcomes = "post-test")
)
```

---

mb_mde *Minimum detectable effect (MDE)*

Description

Inverts mb_power(): given a sample size, target power, and significance level, returns the smallest effect size the design can reliably detect.

Usage

```
mb_mde(
  n_per_group,
  sd = 1,
  power = 0.8,
  alpha = 0.05,
  sides = 2L,
```

```
  type = c("mean", "proportion"),
  baseline = NULL
)
```

Arguments

n_per_group	Numeric. Sample size per arm.
sd	Numeric. Standard deviation, used only for type = "mean". Default 1, in which case effect_size is interpreted in standard deviation units.
power	Numeric in (0, 1). Target power. Default 0.8.
alpha	Numeric in (0, 1). Significance level. Default 0.05.
sides	Integer. 2 (two-sided, default) or 1 (one-sided).
type	Character. "mean" (default) or "proportion".
baseline	Optional numeric in (0, 1). For type = "proportion", the baseline proportion p1 against which the MDE is calculated. The MDE is then returned in absolute proportion-point units.

Value

Numeric scalar. The minimum detectable effect in the units implied by type: standard deviation units (type = "mean", with sd = 1) or absolute proportion-point difference (type = "proportion" with baseline supplied), or Cohen's h (type = "proportion" without baseline).

See Also

mb_power(), mb_sample_size().

Other power: mb_cluster_design(), mb_icc_reference(), mb_power(), mb_sample_size(), mb_stepped_wedge()

Examples

```
mb_mde(n_per_group = 200)
mb_mde(n_per_group = 500, type = "proportion", baseline = 0.4)
```

---

mb_power *Power for a two-sample test*

Description

Computes statistical power for a two-sample test of equal-sized arms, using the large-sample normal approximation. Supports tests of two means (with a common standard deviation) or two proportions (using Cohen's h arcsine effect size).

Usage

```
mb_power(
  n_per_group,
  effect_size = NULL,
  sd = 1,
  alpha = 0.05,
  sides = 2L,
  type = c("mean", "proportion"),
  p1 = NULL,
  p2 = NULL
)
```

Arguments

n_per_group	Numeric. Sample size per arm.
effect_size	Numeric. The standardised effect size: Cohen's d for type = "mean", or Cohen's h for type = "proportion" (computed automatically if p1 and p2 are supplied).
sd	Numeric. Standard deviation, used only for type = "mean". Default 1, in which case effect_size is interpreted in standard deviation units.
alpha	Numeric in (0, 1). Significance level. Default 0.05.
sides	Integer. 2 (two-sided, default) or 1 (one-sided).
type	Character. "mean" (default) or "proportion".
p1, p2	Optional numeric in (0, 1). If both supplied (and type = "proportion"), the function computes Cohen's h and ignores effect_size.

Details

For two means, power is

$$1 - \Phi(z_{1-\alpha/s} - d\sqrt{n/2}) + \Phi(-z_{1-\alpha/s} - d\sqrt{n/2})$$

where s is sides and d is the standardised effect. For two proportions, the effect uses the arcsine variance-stabilising transform: $h = 2\arcsin\sqrt{p_1} - 2\arcsin\sqrt{p_2}$.

Approximation note: this implementation uses the large-sample normal approximation. The standard alternative (used by **pwr**::pwr.t.test) uses the noncentral t-distribution. For typical evaluation sample sizes (n_per_group >= 50) the two agree to within 1-2 percentage points of power; for n_per_group < 30 the discrepancy is larger and **pwr** should be preferred. magentabook ships equivalence tests against **pwr** (see tests/testthat/test-pwr-equivalence.R).

Value

Numeric scalar in (0, 1): the power.

References

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Lawrence Erlbaum.

Champely, S. (2020). pwr: Basic Functions for Power Analysis. R package version 1.3-0. https://CRAN.R-project.org/package=pwr.

HM Treasury (2026). The Magenta Book: Central Government Guidance on Evaluation. Chapter 3 on evaluation methods; further guidance on power analysis in the Transparency in Government Evaluation Research (TIGER) annex. https://www.gov.uk/government/publications/the-magenta-book.

See Also

mb_mde(), mb_sample_size(), mb_cluster_design().

Other power: mb_cluster_design(), mb_icc_reference(), mb_mde(), mb_sample_size(), mb_stepped_wedge()

Examples

```
mb_power(n_per_group = 200, effect_size = 0.3)
mb_power(n_per_group = 500, type = "proportion", p1 = 0.40, p2 = 0.50)
```

---

mb_qaly *Quality-adjusted life years (QALYs) accumulator*

Description

Sums utility-weighted years lived across persons, with optional annual discounting.

Usage

```
mb_qaly(utility, persons = 1, years = 1, discount_rate = NULL)
```

Arguments

utility	Numeric scalar or vector in [0, 1]. Utility weight per year. Length 1 or years.
persons	Numeric scalar. Number of persons. Default 1.
years	Numeric scalar. Number of years. Default 1.
discount_rate	Optional numeric in [0, 1). Annual discount rate. If supplied, returns the discounted QALY total. Default NULL (undiscounted).

Details

Without discounting:

$$\text{QALY} = \text{persons} \cdot \sum_{t=0}^{T-1} u_t$$

With annual discount rate `r`:

$$\text{QALY} = \text{persons} \cdot \sum_{t=0}^{T-1} \frac{u_t}{(1+r)^t}$$

Compatible with `greenbook::gb_qaly`: when `utility` is scalar and `discount_rate` is `NULL`, this returns `persons * utility * years`.

Value

Numeric scalar. Total QALYs.

References

Drummond, M. F. et al. (2015). Methods for the Economic Evaluation of Health Care Programmes (4th ed.). OUP.

NICE (2022). Guide to the methods of technology appraisal.

See Also

`mb_daly()`, `mb_cea()`.

Other cost-effectiveness: `mb_cea()`, `mb_ceac()`, `mb_daly()`, `mb_icer()`, `mb_inb()`

Examples

```
mb_qaly(utility = 0.8, persons = 100, years = 5)
mb_qaly(utility = 0.8, persons = 100, years = 5, discount_rate = 0.035)
mb_qaly(utility = c(0.5, 0.7, 0.9), persons = 50)
```

---

mb_questions *Tag and structure evaluation questions*

Description

Stores a set of evaluation questions tagged by Magenta Book type (process, impact, economic, value-for-money) and by priority (primary or secondary). The Magenta Book canonical taxonomy is bundled in `mb_schedule_table()` under `"questions"`.

Usage

```
mb_questions(text, type = "impact", priority = "primary")
```

Arguments

`text`	Character vector of evaluation questions.
`type`	Character vector. One of `"process"`, `"impact"`, `"economic"`, `"vfm"`. Length 1 or `length(text)`.
`priority`	Character vector. `"primary"` or `"secondary"`. Length 1 or `length(text)`.

Value

An `mb_questions` data frame with columns `text`, `type`, `priority`.

References

HM Treasury (2026). The Magenta Book: Central Government Guidance on Evaluation. Chapter 1.8 on types of evaluation (process, impact, value for money); Chapter 2 on evaluation scoping. https://www.gov.uk/government/publications/the-magenta-book.

See Also

`mb_evaluation_plan()`, `mb_schedule_table()`.

Other planning: `mb_balance_table()`, `mb_counterfactual()`, `mb_evaluation_plan()`, `mb_stakeholders()`

Examples

```
mb_questions(
  text     = c("Did the policy cause employment to rise",
               "Was implementation faithful to the design"),
  type     = c("impact", "process"),
  priority = c("primary", "secondary")
)
```

---

mb_sample_size *Required sample size for a target power*

Description

Given a target effect size, power, and significance level, returns the required sample size per arm. Inverts `mb_power()`.

Usage

```
mb_sample_size(
  effect_size = NULL,
  sd = 1,
  power = 0.8,
  alpha = 0.05,
  sides = 2L,
  type = c("mean", "proportion"),
```

```
  p1 = NULL,
  p2 = NULL
)
```

Arguments

effect_size	Numeric. The standardised effect size: Cohen's d for type = "mean", or Cohen's h for type = "proportion" (computed automatically if p1 and p2 are supplied).
sd	Numeric. Standard deviation, used only for type = "mean". Default 1, in which case effect_size is interpreted in standard deviation units.
power	Numeric in (0, 1). Target power. Default 0.8.
alpha	Numeric in (0, 1). Significance level. Default 0.05.
sides	Integer. 2 (two-sided, default) or 1 (one-sided).
type	Character. "mean" (default) or "proportion".
p1, p2	Optional numeric in (0, 1). If both supplied (and type = "proportion"), the function computes Cohen's h and ignores effect_size.

Value

Integer scalar. Sample size per arm (rounded up).

See Also

mb_power(), mb_mde(), mb_cluster_design().

Other power: mb_cluster_design(), mb_icc_reference(), mb_mde(), mb_power(), mb_stepped_wedge()

Examples

```
mb_sample_size(effect_size = 0.3, power = 0.8)
mb_sample_size(type = "proportion", p1 = 0.40, p2 = 0.50, power = 0.8)
```

---

mb_schedule_table *Expose internal lookup tables*

Description

Returns one of the bundled lookup tables: the Maryland SMS rubric, the Magenta Book confidence rubric, the ICC reference table, or the evaluation question taxonomy.

Usage

```
mb_schedule_table(table = c("sms", "confidence", "icc", "questions"))
```

Arguments

table	Character scalar. One of "sms", "confidence", "icc", or "questions".

Value

A data frame.

See Also

`mb_data_versions()`.

Other lookups: `mb_data_versions()`

Examples

```
mb_schedule_table("sms")
mb_schedule_table("confidence")
mb_schedule_table("icc")
mb_schedule_table("questions")
```

---

mb_sms_explain *Explain the Maryland SMS rubric*

Description

Prints the bundled Maryland SMS rubric. Use this when scoring studies, training reviewers, or presenting evidence ratings to stakeholders.

Usage

```
mb_sms_explain(level = NULL)
```

Arguments

`level`	Optional integer in `1:5`. If supplied, prints the rubric for that single level. If `NULL` (default), prints the full rubric.

Value

Invisibly, the rubric data frame (filtered to `level` if supplied). Called for the side-effect of printing.

See Also

`mb_sms_rate()`.

Other Maryland SMS: `mb_sms_rate()`

Examples

```
mb_sms_explain()
mb_sms_explain(4)
```

`mb_sms_rate` *Score a study against the Maryland Scientific Methods Scale*

Description

Records an evidence rating against the 1-5 Maryland SMS, the What Works Network's standard for grading impact evidence.

Usage

```
mb_sms_rate(level, study, design = NULL, notes = NULL)
```

Arguments

`level`	Integer in `1:5`. The Maryland SMS level.
`study`	Character scalar. Reference for the study being rated (citation, URL, internal ID).
`design`	Optional character scalar. Brief description of the design (e.g. `"Cluster RCT, 80 schools"`).
`notes`	Optional character scalar. Additional notes on methodological strengths and weaknesses.

Details

The Maryland SMS, originally developed by Sherman et al. (1997) for crime-prevention research, is the foundation for evidence ratings used by the College of Policing What Works Centre, the Education Endowment Foundation, the Early Intervention Foundation, and others. The Magenta Book adopts SMS as its default for grading impact evidence.

Level 1: cross-sectional or before-after with no comparison. Level 2: before-after with a non-equivalent comparison group. Level 3: well-matched comparison across multiple units. Level 4: comparison adjusting for unobservables (DiD, RD, IV, ITS, synthetic control). Level 5: random assignment.

Provenance note: numeric levels 1-5 are direct from Sherman et al. (1997). The word labels (Weakest / Weak / Moderate / Strong / Strongest) follow What Works UK / Education Endowment Foundation convention and are not direct quotations from the original report. The design-examples and typical-use columns of the bundled rubric are magentabook synthesis, intended as a practitioner reference rather than a verbatim reproduction.

Value

An `mb_sms_rating` object: a list capturing the level, study, design, notes, the corresponding rubric row, and `vintage`.

References

Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P., Bushway, S. (1997). Preventing Crime: What Works, What Doesn't, What's Promising. Report to the US Congress. Original Maryland Scientific Methods Scale.

The Maryland Scientific Methods Scale is not named explicitly in the May 2026 republication of the Magenta Book, but the underlying hierarchy of evaluation rigour (Sherman et al., 1997) remains widely used across UK What Works Centres (Education Endowment Foundation, College of Policing, Justice Data Lab, Early Intervention Foundation) for rating quasi-experimental designs. The 2026 edition discusses general method selection in Chapter 3 on evaluation methods and in Annex A. https://www.gov.uk/government/publications/the-magenta-book.

See Also

`mb_sms_explain()`, `mb_confidence()`.

Other Maryland SMS: `mb_sms_explain()`

Examples

```
mb_sms_rate(
  level  = 5,
  study  = "Card & Krueger (1994) NJ minimum wage",
  design = "Difference-in-differences with PA comparison",
  notes  = "Large N, but contested measurement"
)
```

---

mb_stakeholders *RACI-style stakeholder register*

Description

Records who is Responsible, Accountable, Consulted, or Informed for an evaluation, with optional interest and influence ratings for use in a stakeholder map.

Usage

```
mb_stakeholders(name, role, raci, interest = NA_real_, influence = NA_real_)
```

Arguments

`name`	Character vector of stakeholder names.
`role`	Character vector of stakeholder roles.
`raci`	Character vector. One of `"R"`, `"A"`, `"C"`, `"I"`.
`interest`	Optional numeric vector in `[1, 5]`. Higher means more interest in the evaluation.
`influence`	Optional numeric vector in `[1, 5]`. Higher means more influence over the evaluation.

Value

An `mb_stakeholders` data frame with columns `name`, `role`, `raci`, `interest`, `influence`.

See Also

`mb_evaluation_plan()`.

Other planning: `mb_balance_table()`, `mb_counterfactual()`, `mb_evaluation_plan()`, `mb_questions()`

Examples

```
mb_stakeholders(
  name      = c("HMT", "DfE", "What Works Centre"),
  role      = c("Funder", "Delivery", "Synthesis"),
  raci      = c("A", "R", "C"),
  interest  = c(5, 5, 4),
  influence = c(5, 4, 2)
)
```

---

mb_stepped_wedge *Stepped-wedge design effect*

Description

Computes the design effect for a stepped-wedge cluster randomised trial relative to an individually-randomised parallel design with the same total observations.

Usage

```
mb_stepped_wedge(steps, clusters_per_step, individuals_per_cluster, icc)
```

Arguments

`steps`	Integer. Number of measurement periods (also called `T`). Includes the baseline.
`clusters_per_step`	Numeric. Number of clusters that crossover at each step.
`individuals_per_cluster`	Numeric. Individuals measured per cluster per period.
`icc`	Numeric in `[0, 1]`. Intra-class correlation coefficient.

Details

Implements the closed-form approximation from Hemming et al. (2015) BMJ Box 2:

Within-cluster design effect (cluster RCT vs individual RCT with same total observations):

$$\text{DEFF}_c = 1 + (mT - 1)\rho$$

Stepped-wedge correction relative to a parallel cluster RCT:

$$\mathrm{CF} = \frac{3(1-\rho)}{2T(1-1/T^2)}$$

Combined: `DEFF_sw = DEFF_c * CF`. This is a multiplier on the variance of the treatment effect compared with an individually-randomised design with the same total observations.

Approximation note: this is the closed-form approximation. The exact Hussey-Hughes (2007) variance, which `swCRTdesign::swPwr` computes from the design matrix, can differ by 20-40 percent for typical UK evaluation designs. magentabook ships a cross-validation test (`tests/testthat/test-swcrt-equivalence.R`) that documents the magnitude of this approximation gap on a grid of designs. For production sample-size work, especially where rho is high or the number of steps is small, prefer `swCRTdesign::swPwr` or `clusterPower::cps.sw.binary` over this function. Use `mb_stepped_wedge` for quick comparative exploration; use the specialist packages for the number you commit to in a published evaluation plan.

Both forms assume a balanced design: equal cluster size, equal-period intervals, complete data, no time-by-treatment interaction, and one outcome measurement per cluster-period. For non-standard designs use the specialist packages above.

Value

A list with elements `deff_cluster` (the within-period cluster design effect), `correction_factor` (the stepped-wedge correction relative to a parallel cluster RCT), `deff_sw` (the product), and `n_total` (total observations across the trial).

References

Hussey, M. A., Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. Contemporary Clinical Trials 28(2). doi:10.1016/j.cct.2006.05.007.

Woertman, W., de Hoop, E., Moerbeek, M., Zuidema, S. U., Gerritsen, D. L., Teerenstra, S. (2013). Stepped wedge designs could reduce the required sample size in cluster randomized trials. Journal of Clinical Epidemiology 66(7). doi:10.1016/j.jclinepi.2012.12.003.

Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J., Lilford, R. J. (2015). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ 350. doi:10.1136/bmj.h391.

See Also

`mb_cluster_design()`, `mb_icc_reference()`.

Other power: `mb_cluster_design()`, `mb_icc_reference()`, `mb_mde()`, `mb_power()`, `mb_sample_size()`

Examples

```
mb_stepped_wedge(
  steps = 5,
  clusters_per_step = 4,
  individuals_per_cluster = 20,
  icc = 0.05
)
```

mb_theory_of_change *Build a Magenta Book theory of change*

Description

Constructs a five-level logic model in the form set out by the HM Treasury Magenta Book: inputs → activities → outputs → outcomes → impact, with optional assumptions and external factors.

Usage

```
mb_theory_of_change(
  inputs,
  activities,
  outputs,
  outcomes,
  impact,
  assumptions = NULL,
  external_factors = NULL,
  name = NULL
)
```

Arguments

inputs	Character vector of resources committed to the policy: funding, staff, infrastructure, partnerships.
activities	Character vector of what the policy does with those inputs: design, delivery, communication, enforcement.
outputs	Character vector of direct, countable products of the activities: training sessions delivered, leaflets posted, payments made.
outcomes	Character vector of changes the outputs produce in the target population, typically over months to a few years: behaviour change, attitudes, take-up.
impact	Character vector of long-term, ultimate goals the outcomes contribute to: poverty reduction, decarbonisation, improved health.
assumptions	Optional character vector of assumptions that must hold for each level to translate into the next.
external_factors	Optional character vector of contextual factors outside the policy's control that may affect outcomes.
name	Optional character scalar naming the policy or programme.

Details

The Magenta Book theory of change is the foundation for every subsequent evaluation step. It makes the implicit causal chain explicit so that evaluation questions can be tied to specific levels and indicators can be defined.

Value

An mb_toc object: a list with one element per level plus optional assumptions, external_factors, name, and vintage.

References

HM Treasury (2026). The Magenta Book: Central Government Guidance on Evaluation. Chapter 2 on evaluation scoping; Annex A Section A1 on theory-based methods for impact evaluation. https://www.gov.uk/government/publications/the-magenta-book.

See Also

mb_logframe(), mb_assumptions().

Other theory of change: mb_assumptions(), mb_logframe()

Examples

```
toc <- mb_theory_of_change(
  inputs     = c("GBP 50m grant", "12 FTE programme team"),
  activities = c("Design training", "Deliver workshops"),
  outputs    = c("500 workshops delivered", "8000 attendees"),
  outcomes   = c("Improved skills", "Increased confidence"),
  impact     = "Higher employment among target group",
  assumptions = "Workshops cause skills uplift",
  external_factors = "Macro labour market remains stable",
  name = "Skills uplift programme"
)
toc
```

---

mb_to_excel *Export an evaluation report to Excel*

Description

Writes a multi-sheet workbook with one sheet per component: summary, theory of change, plan, SMS ratings, confidence ratings, cost-effectiveness, provenance.

Usage

```
mb_to_excel(report, file)
```

Arguments

report	An mb_report object.
file	Output file path (must end in .xlsx).

Details

Requires the **openxlsx** package (in Suggests).

Value

Invisibly, the file path.

See Also

`mb_to_word()`, `mb_to_latex()`.

Other reporting: `mb_evaluation_report()`, `mb_to_latex()`, `mb_to_word()`

Examples

```
if (requireNamespace("openxlsx", quietly = TRUE)) {
  toc <- mb_theory_of_change(
    inputs = "Funding", activities = "Workshops",
    outputs = "Attendees", outcomes = "Skills",
    impact = "Employment"
  )
  rep <- mb_evaluation_report(toc = toc, name = "Skills uplift")
  tmp <- tempfile(fileext = ".xlsx")
  mb_to_excel(rep, tmp)
}
```

---

mb_to_latex *Render an evaluation report as a LaTeX table*

Description

Returns a single LaTeX `tabular` summarising the report. Multi-sheet Word/Excel exports are richer; LaTeX is intended for insertion into a one-pager.

Usage

```
mb_to_latex(report, caption = NULL, label = NULL)
```

Arguments

`report`	An `mb_report` object.
`caption`	Optional table caption.
`label`	Optional LaTeX label for cross-referencing.

Value

A character scalar containing a LaTeX `tabular` environment.

See Also

mb_to_word(), mb_to_excel().

Other reporting: mb_evaluation_report(), mb_to_excel(), mb_to_word()

Examples

```
toc <- mb_theory_of_change(
  inputs = "Funding", activities = "Workshops",
  outputs = "Attendees", outcomes = "Skills",
  impact = "Employment"
)
rep <- mb_evaluation_report(toc = toc, name = "Skills uplift")
cat(mb_to_latex(rep))
```

---

mb_to_word *Export an evaluation report to Word*

Description

Writes a one- to two-page Word document summarising an mb_report: name, theory of change, evaluation plan, SMS ratings, confidence ratings, and cost-effectiveness.

Usage

```
mb_to_word(report, file)
```

Arguments

report	An mb_report object.
file	Output file path (must end in .docx).

Details

Requires the **officer** and **flextable** packages (both in Suggests).

Value

Invisibly, the file path.

See Also

mb_evaluation_report(), mb_to_excel(), mb_to_latex().

Other reporting: mb_evaluation_report(), mb_to_excel(), mb_to_latex()

Examples

```
if (requireNamespace("officer", quietly = TRUE) &&
    requireNamespace("flextable", quietly = TRUE)) {
  toc <- mb_theory_of_change(
    inputs = "Funding", activities = "Workshops",
    outputs = "Attendees", outcomes = "Skills",
    impact = "Employment"
  )
  rep <- mb_evaluation_report(toc = toc, name = "Skills uplift")
  tmp <- tempfile(fileext = ".docx")
  mb_to_word(rep, tmp)
}
```

Index

∗ **Maryland SMS**
  mb_sms_explain, 34
  mb_sms_rate, 35
∗ **confidence**
  mb_confidence, 9
  mb_confidence_summary, 11
∗ **cost-effectiveness**
  mb_cea, 5
  mb_ceac, 6
  mb_daly, 14
  mb_icer, 23
  mb_inb, 24
  mb_qaly, 30
∗ **estimators**
  mb_did_2x2, 16
  mb_event_study, 20
  mb_its, 25
∗ **lookups**
  mb_data_versions, 15
  mb_schedule_table, 33
∗ **planning**
  mb_balance_table, 4
  mb_counterfactual, 13
  mb_evaluation_plan, 17
  mb_questions, 31
  mb_stakeholders, 36
∗ **power**
  mb_cluster_design, 7
  mb_icc_reference, 22
  mb_mde, 27
  mb_power, 28
  mb_sample_size, 32
  mb_stepped_wedge, 37
∗ **realist**
  mb_cmo, 8
  mb_contribution_claim, 12
∗ **reporting**
  mb_evaluation_report, 19
  mb_to_excel, 40
  mb_to_latex, 41
  mb_to_word, 42
∗ **theory of change**
  mb_assumptions, 3
  mb_logframe, 26
  mb_theory_of_change, 39

mb_assumptions, 3, *27*, *40*
mb_assumptions(), *40*
mb_balance_table, 4, *14*, *18*, *32*, *37*
mb_cea, 5, *7*, *15*, *24, 25*, *31*
mb_cea(), *24*, *31*
mb_ceac, *6*, 6, *15*, *24, 25*, *31*
mb_ceac(), *6*, *24, 25*
mb_cluster_design, 7, *23*, *28*, *30*, *33*, *38*
mb_cluster_design(), *23*, *30*, *33*, *38*
mb_cmo, 8, *13*
mb_cmo(), *13*
mb_confidence, 9, *11*
mb_confidence(), *11*, *36*
mb_confidence_summary, *11*, 11
mb_confidence_summary(), *11*
mb_contribution_claim, *9*, 12
mb_contribution_claim(), *9*
mb_counterfactual, *5*, 13, *18*, *32*, *37*
mb_counterfactual(), *18*
mb_daly, *6, 7*, 14, *24, 25*, *31*
mb_daly(), *31*
mb_data_versions, 15, *34*
mb_data_versions(), *34*
mb_did_2x2, 16, *21*, *26*
mb_did_2x2(), *5*, *21*, *26*
mb_evaluation_plan, *5*, *14*, 17, *32*, *37*
mb_evaluation_plan(), *14*, *19*, *32*, *37*
mb_evaluation_report, 19, *41, 42*
mb_evaluation_report(), *18*, *42*
mb_event_study, *17*, 20, *26*
mb_event_study(), *17*, *26*
mb_icc_reference, *8*, 22, *28*, *30*, *33*, *38*
mb_icc_reference(), *7, 8*, *38*

mb_icer, *6, 7*, *15*, 23, *25*, *31*
mb_icer(), *5–7*, *25*
mb_inb, *6, 7*, *15*, *24*, 24, *31*
mb_inb(), *6, 7*, *24*
mb_its, *17*, *21*, 25
mb_its(), *17*, *21*
mb_logframe, *3*, 26, *40*
mb_logframe(), *3*, *40*
mb_mde, *8*, *23*, 27, *30*, *33*, *38*
mb_mde(), *30*, *33*
mb_power, *8*, *23*, *28*, 28, *33*, *38*
mb_power(), *27, 28*, *32, 33*
mb_qaly, *6, 7*, *15*, *24, 25*, 30
mb_qaly(), *15*
mb_questions, *5*, *14*, *18*, 31, *37*
mb_questions(), *5*, *18*
mb_sample_size, *8*, *23*, *28*, *30*, 32, *38*
mb_sample_size(), *8*, *28*, *30*
mb_schedule_table, *16*, 33
mb_schedule_table(), *10*, *16*, *31, 32*
mb_sms_explain, 34, *36*
mb_sms_explain(), *36*
mb_sms_rate, *34*, 35
mb_sms_rate(), *11*, *34*
mb_stakeholders, *5*, *14*, *18*, *32*, 36
mb_stakeholders(), *18*
mb_stepped_wedge, *8*, *23*, *28*, *30*, *33*, 37
mb_stepped_wedge(), *8*, *23*
mb_theory_of_change, *3*, *27*, 39
mb_theory_of_change(), *3*, *19*, *27*
mb_to_excel, *19*, 40, *42*
mb_to_excel(), *19*, *42*
mb_to_latex, *19*, *41*, 41, *42*
mb_to_latex(), *19*, *41, 42*
mb_to_word, *19*, *41, 42*, 42
mb_to_word(), *19*, *41, 42*

pwr.t.test, *29*