Effect Sizes & Statistical Reporting in R

Match the estimate to the design; pair every effect with uncertainty

Rverse Analytics

A *p*-value answers compatibility under a null model; an effect size answers magnitude. Prefer a scale with clinical meaning, then add a standardized measure only when it improves comparison.

Effect-size map

Design / outcome	Primary effect	R recipe	Interpretation anchor
Two independent means	mean difference	`t.test(y ~ group, data = d)`	outcome units
Standardized means	Hedges' *g*	`effectsize::hedges_g(y ~ group, data = d)`	pooled SD units
Paired continuous	mean change	`t.test(post, pre, paired = TRUE)`	within-person units
≥3 means	η² / ω²	`effectsize::omega_squared(aov_fit)`	variance proportion
Rank comparison	rank-biserial *r*	`effectsize::rank_biserial(y ~ group, data = d)`	dominance on [−1, 1]
Two continuous	Pearson/Spearman *r*	`cor.test(x, y, method = "spearman")`	association, not agreement
Binary outcome	RD, RR, OR	formulas / regression	absolute and relative risk
Count rate	incidence-rate ratio	Poisson model + offset	event-rate ratio
Time-to-event	HR + survival difference	Cox + `survfit()`	relative hazard + absolute S(t)
Prediction	calibration + discrimination	calibration plot + AUC	reliability + ranking

Continuous outcomes

```
library(effectsize)

# Raw difference and CI remain the most interpretable
t.test(y ~ group, data = d, var.equal = FALSE)

# Bias-corrected standardized difference
hedges_g(y ~ group, data = d, pooled_sd = TRUE, ci = 0.95)

# Omnibus ANOVA effect; omega squared is less biased than eta squared
fit_aov <- aov(y ~ group, data = d)
omega_squared(fit_aov, partial = FALSE, ci = 0.95)

# Paired design: state the standardizer used
repeated_measures_d(d$post, d$pre, method = "z", ci = 0.95)
```

Do not label 0.2/0.5/0.8 automatically "small/medium/large." Context, outcome reliability, baseline risk and clinical thresholds determine importance.

Binary outcomes: show absolute and relative effects

```
# Exposed/treatment risk and control risk
r1 <- events_treat / n_treat
r0 <- events_control / n_control

rd  <- r1 - r0
rr  <- r1 / r0
orr <- (events_treat / (n_treat - events_treat)) /
       (events_control / (n_control - events_control))

# NNT is horizon- and outcome-specific; use absolute risk difference
nnt <- 1 / abs(rd)

# Adjusted odds ratios
fit <- glm(event ~ treatment + age + severity,
           family = binomial, data = d)
broom::tidy(fit, exponentiate = TRUE, conf.int = TRUE)
```

Quantity	Strength	Common misuse
Risk difference	clinical absolute impact	transported across very different baseline risks
Risk ratio	intuitive relative effect	called an odds ratio
Odds ratio	natural for logistic/case-control analysis	interpreted as RR when outcome is common
NNT / NNH	decision-friendly at fixed horizon	reported without time, CI or event direction

Confidence intervals and multiplicity

```r
# Bootstrap a statistic when analytic assumptions are doubtful
set.seed(20260711)
B <- replicate(2000, {
  idx <- sample.int(nrow(d), replace = TRUE)
  mean(d$y[idx], na.rm = TRUE)
})
quantile(B, c(.025, .975), na.rm = TRUE)

# Adjust a family of p-values; keep raw estimates and CIs visible
p.adjust(p_values, method = "BH")         # false-discovery rate
p.adjust(p_values, method = "holm")       # family-wise error
```

Resample at the independent-unit level: patients for ordinary studies, clusters for cluster trials, and subjects (not rows) for repeated measures.

APA-style reporting patterns

Analysis	Compact pattern
Welch *t*	“Groups differed by 4.2 points, 95% CI [1.1, 7.3], $t(73.4) = 2.70$, $p = .009$, Hedges’ $g = 0.61$.”
ANOVA	“The group effect was significant, $F(2, 87) = 5.42$, $p = .006$, $\omega^2 = .09$, 95% CI [.01, .18].”
Correlation	“Scores were positively associated, $r(98) = .34$, 95% CI [.15, .51], $p < .001$.”
Logistic	“Treatment was associated with lower odds, adjusted OR = 0.62, 95% CI [0.41, 0.94], $p = .024$.”
Cox	“The adjusted hazard was lower, HR = 0.70, 95% CI [0.52, 0.95], $p = .021$.”

Final audit

- Identify the effect, direction, unit, reference group, analysis population and time horizon.
- Keep estimates and confidence intervals at higher prominence than significance stars.
- Use enough digits to preserve meaning; do not imply precision the design cannot support.
- Report assumption checks and sensitivity analyses that can change the conclusion.
- Distinguish statistical uncertainty from clinical importance and generalizability.
