Reporting Statistical Results in Text and in Graphs

The Office of Evaluation Sciences reports evaluation results in the form of abstracts published on its website. Most evaluations are done via randomized controlled trials (RCTs), and results typically are presented for a linear regression model estimated with ordinary least squares (OLS) where one of the regression model's parameters represents the effect of a treatment or intervention. This document describes OES's preferred method of reporting these estimates, 95% confidence intervals, and *p*-values.

When to Apply This Guidance

When reporting treatment effects and statistical inferences, OES has a preferred style for reporting results in text and for graphical presentation. These guidelines apply both for project abstracts (published on the OES website) and for slide presentations. This guidance pertains most directly to results from RCTs, and observational quasi-experimental designs (QEDs) may require alternative ways of presenting results.

Required Elements in Text

One treatment condition and a control condition

For reporting the treatment effect estimate from a linear regression, required elements to include in text are:

- The observed mean (proportion for binary responses) for the control group
- The point estimate of the regression coefficient (expressed in percentage points for binary responses). Depending on the research design and statistical model, this may be a covariate-adjusted estimate.
- The 95% confidence interval for the regression coefficient
- The *p*-value, rounded to 2 or 3 decimal places as appropriate. Report the actual *p*-value, not just whether it falls below a threshold.
- If multiple comparisons adjustment is used for *p*-values, report the *un*adjusted values in text and note the adjusted values in a footnote (read more about OES guidance on multiple comparisons).

These values help to convey to an audience both the *statistical* and *practical* significance of treatment effects — with the former captured by *p*-values and the latter captured by point estimates and confidence intervals.

Here's an example, from a fictional project that assessed the effect of letters on applications for a benefit (required elements are highlighted in **bold**):

Of individuals who received a letter, 6.00 percent applied for benefits in the nine months after the letters were sent out, compared with **0.96 percent of individuals who did not receive a letter**. After adjusting for age, socioeconomic status, and other factors that might influence whether an individual applies for benefits, we estimated that letters resulted in **an increase of 5.09 percentage points (p =.000, 95% CI [5.02, 5.17])**.

Note two things about how the treatment effect is described in this example. First, the text makes clear that the estimated treatment effect (an increase of 5.09 percentage points) is not just the difference between the observed means in the two experimental groups but rather is adjusted for covariates included in the statistical model. Second, the text is clear that this treatment effect on a binary outcome is measured in percentage points (as opposed to "an increase of 5.09 percent"). Effects on binary outcomes should be described in percentage points. If it is also useful to describe the effect in relative terms as a percentage of the baseline outcome ("a 530% increase"), then this can be done too, but should be done in addition to, not instead of, an estimate in percentage points.

Multiple treatment conditions and a control condition

Many evaluations include two or more treatment conditions. When this is the case, the preferred elements to include in text are:

- The observed mean (proportion for binary responses) for the control group
- The point estimate (regression coefficient) for the treatment effect for each treatment group, relative to the control/reference group. Depending on the research design and statistical model, these may be covariate-adjusted treatment effects.
- The 95% confidence intervals for these point estimates
- The p-values for these treatment effects, rounded to 2 or 3 decimal places as appropriate. Report the actual p-values, not just whether they fall below a threshold.
- If adjusting p-values for multiple comparisons, report the *un*adjusted values in text and note the adjusted values in a footnote (read more about OES guidance on multiple comparisons).

Depending on the research questions and comparisons specified in the analysis plan, differences among the treatment arms may be relevant. In each case, the (potentially covariate-adjusted) difference should be reported as described above (point estimate of the difference, 95% CI, and actual p-value).

Predicted means for the treatment arms may also be explicitly reported if relevant, but this is optional (these are likely depicted graphically; see guidance on figures below).

Here's an example from project 1738, which evaluated the effect of postcard reminders to seniors on vaccine uptake. The trial involved a stepped-wedge design where timing of receiving postcards

was varied for those assigned to treatment for a total of three treatment arms and one control group. The results were summarized in text as follows:[1]

> **Individuals in the control group received 8.58 percent** of recommended vaccinations during the study period. Results indicated that, compared to the control group, individuals in the October group showed **an increase of 0.27 percentage points (p = .002, CI [0.12, 0.42])**, and individuals in the November group showed **an increase of 0.15 percentage points (p = .13, CI [0.00, 0.30])**.

In this particular example, the estimated treatment effect was statistically significant for only one of the treatment arms. When this is the case, it may be useful to highlight this in-text. For example: "The treatment effect was statistically significant only for the October group."

Required Elements in Figures

In addition to reporting results in-text, it is helpful to present results graphically. This gives readers a frame of reference for interpreting the magnitude of causal effects relative to the baseline response.

One treatment condition and a control condition

When there is one treatment condition, required elements for graphical presentation are:

- Figure should be in the form of a bar plot where the y-axis starts at zero;
- Report the *observed* mean for the control group with a blue bar;
- Report the (potentially covariate-adjusted) *predicted* mean for the treatment group in yellow (control mean + treatment effect);[2]
- Include 95% confidence intervals for those means. For the control group, the 95% confidence interval is based on the standard error of the mean. For the treatment group, the 95% confidence interval is based on the standard error of the treatment effect estimate.
- In a footnote, explain (1) that the treatment group estimate is based on our estimate of the treatment effect and covariate adjusted (if applicable); (2) that the error bars represent 95% confidence intervals; and (3) that the confidence interval for the treatment condition reflects uncertainty in our treatment effect estimate and covariate adjustment (if applicable). The last point is important to make clear that the confidence intervals for the two groups are calculated in different ways and represent different sources of uncertainty;
    - Here is an example to be adapted as appropriate: "The estimate for the email group is based on our estimate of the treatment effect after adjusting for age, race/ethnicity, and other covariates. Error bars represent 95% confidence intervals, and the confidence interval for the email group reflects statistical

---

[1] This example has been edited for clarity.
[2] For more on what to report when using blocking, see OES guidance on block randomization.

uncertainty in our estimate of the treatment effect after adjustment for covariates."

- Optional markers of statistical significance (e.g., an asterisk), with a legend indicating the *p*-value threshold for statistical significance;
- A horizontal line denoting the observed mean for the control condition is strongly recommended in most cases, but not necessary;
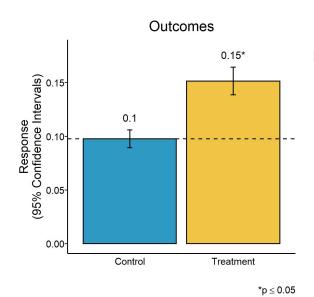- Include a y-axis label indicating the scale and type of response (e.g., "proportion applications").

Figure 1: *Example figure using "dummy" data.[3]*

Figure 1 shows an example using some simulated data. It includes the observed mean of the control group (blue) and the predicted mean of the treatment group (yellow). It also includes 95% confidence intervals and further includes a horizontal line at the level of the control mean to clarify that the significance of the treatment effect is evaluated with respect to the response under treatment relative to the mean under control. The y-axis further begins at zero, and the y-axis label notes that the intervals shown are 95% confidence intervals.

Multiple treatment conditions and a control condition

When there are two or more treatment conditions:

- Figure should be in the form of a bar plot where the y-axis starts at zero;
- Report the *observed* mean for the control group with a blue bar;
- Report the (potentially covariate-adjusted) *predicted* means for the treatment groups in yellow (control mean + treatment effect);[4]

---

[3] R code to produce this figure is included at the end of this document.
[4] For more on what to report when using blocking, see OES guidance on block randomization.

- Include 95% confidence intervals for those means. For the control group, the 95% confidence interval is based on the standard error of the mean. For the treatment groups, the 95% confidence intervals are based on the standard error of the treatment effect estimates.
- In a footnote, explain (1) that the treatment group estimates are based on our estimates of the treatment effects per treatment arm and covariate adjusted (if applicable); (2) that the error bars represent 95% confidence intervals; and (3) that the confidence interval for each treatment condition reflects uncertainty in our treatment effect estimate and covariate adjustment (if applicable). The last point is important to make clear that the confidence intervals for the two groups are calculated in different ways and represent different sources of uncertainty;
    - Here is an example to be adapted as appropriate: "The predicted means for the information-only email group and the risk-framing email group are based on our estimates of the treatment effects after adjusting for age, race/ethnicity, and other covariates. Error bars represent 95% confidence intervals, and the confidence interval for each email group reflects statistical uncertainty in our estimate of the treatment effect after adjustment for covariates."
- Optional markers of statistical significance (e.g., an asterisk), with a legend indicating the *p*-value threshold for statistical significance;
- A horizontal line denoting the observed mean for the control condition is strongly recommended in most cases, but not necessary;
- Include a y-axis label indicating the scale and type of response (e.g., "proportion applications").
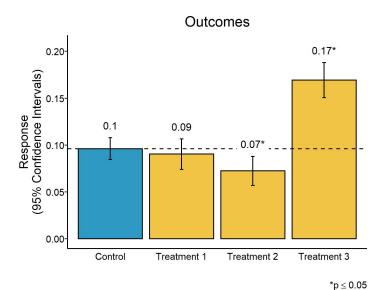

Figure 2: *Example figure using "dummy" data.*[5]

---

[5] R code to produce this figure is included at the end of this document.

Figure 2 provides an example using simulated data. This figure contains all of the recommended elements: a bar plot starting at zero on the y-axis, the control group mean, the predicted treatment means, and 95% confidence intervals.

Code Appendix

Sample code is provided in both R and Stata:

R Code

The below code was used to generate Figures 2 and 4. The source code for the `oes_plot(...)` function can be [found here](#) (this link might not be accessible from outside OES). To use the code, just click "Raw" to view the raw script file and then copy and paste the URL where the placeholder URL is below.

```
# pull oes_plot function from source code
gitHub.url <- "https://Place.Holder…"
source(gitHub.url)

# load and attach estimatr library
library(estimatr)

# df : a data frame of simulated data
# x1 : binary indicator for treatment assignment (two arms)
# x2 : factor indicating assignment to multi-arm treatment (four
arms)
# y1 : response for binary treatment
# y2 : response for multi-arm treatment

# fit models
fit1 <- lm_robust(y1 ~ x1, df)
fit2 <- lm_robust(y2 ~ x2, df)

# plot results for each (figures 2 and 4 in document)
oes_plot(fit1,treatment_arms = 1,title = "Outcomes")
oes_plot(fit2,treatment_arms = 3,title = "Outcomes")
```

Stata Code

The below code was used to generate Figure 4. The source code for the `oes_plot_stata` program can be [found here](#) (this link might not be accessible from outside OES). To run a program in Stata, you can save the GitHub file locally (either pull the file from GitHub or copy and paste it

into a local do file). Then load the program as per the "configure program" line below and run it with any arguments specific to your dataset.

```
/*Note that to maintain consistency, this example code uses the
simulated dataset referenced in the R code but renames x2 ->
treatment and y2 -> y given that the Stata code does not
differentiate between datasets with one or more treatments.*/

//generate dummies for each treatment group
forval i = 0/3 {
    gen treat_`i' = (treatment == `i')
}

//run specification from which to estimate core treatment effects
reg y treat_1 treat_2 treat_3, vce(hc2)

//capture standard errors
forval i = 1/3 {
    local treat_`i'_se = _se[treat_`i']
}

//capture p-values
forval i = 1/3 {
    lincom _b[treat_`i']
    local p_`i' = (2 * ttail(e(df_r), abs(r(estimate)/r(se))))
}

//graph arguments
local ytitle = "Response" // y axis label
local alpha_level = .05 // pvalue cutoff
local ymin = 0 //minimum y axis value (strongly recommend starting at
0!)
local ymax = .15 //maximum y axis value
local ygap = .05 //gap between each y axis tick mark

//configure program
do ".../GitHub/code_library/data_visualization/oes_plot_stata.do"

//run figure - Note: arguments should always appear in the order they
are listed in the program
oes_plot_stata "`ytitle'" "`alpha_level'" "`ymin'" "`ymax'" "`ygap'"
"`p_1'" "`treat_1_se'" "`p_2'" "`treat_2_se'" "`p_3'" "`treat_3_se'"

//export your graph locally!
```