

Errata

On Coppock, Leeper and Mullinix (2018), *Proceedings of the National Academy of Sciences* 115(49)

This note lists 16 corrections to “Generalizability of heterogeneous treatment effect estimates across samples,” published in the *Proceedings of the National Academy of Sciences* 115(49):12441–12446, doi [10.1073/pnas.1808083115](https://doi.org/10.1073/pnas.1808083115). Three are sentences that describe a column of a table incorrectly: two in the supplementary appendix, one in the main text. Five correct a label inside Table 1, Fig. 1 or Fig. 2. The remaining eight are entries of the reference list. **None of them changes any conclusion of the paper**, and none touches an estimate: every CATE, standard error, p-value and confidence limit in the article and in all 27 appendix tables reproduces exactly from the deposited archive.

The numbers in the corrected sentences and every number in the notes beneath them are computed from the deposited replication archive ([10.7910/DVN/4WNGEJ](https://doi.org/10.7910/DVN/4WNGEJ)) at the time this document is rendered. The code and output supporting the correction are at github.com/active-maintenance-aec/coppock_leeper_mullinix_2018.

1. The appendix’s Prop column is computed with the wrong divisor

0.18 for respondents aged 18 to 39 in appendix Table 1’s Original panel

Corrected:

0.35 for respondents aged 18 to 39 in appendix Table 1’s Original panel

The divisor is two in 23 of the 27 study pairs, whose subgroup regressions carry a treatment indicator and an intercept and nothing else, and three, four or five in the remaining four, whose regressions carry a further arm of the study’s own factorial design or the party identification its assignment probabilities depend on. An attribute’s classes therefore sum to between 0.20 and 0.50 across the tables rather than to 1.00. In appendix Table 1, for instance, the Original panel prints 0.18 for respondents aged 18 to 39, where they are in fact 0.35 of that sample.

The error is in the deposited code rather than in the typesetting: the expression that builds the column divides each subgroup’s count by a total taken over every row of the fitted

results rather than over the treatment coefficients alone, and all 787 printed values reproduce exactly under that expression.

This entry used to restate what the column contains rather than correct it, on the ground that the alternative was reprinting 27 tables. Alex, 2026-08-12: “surely N was meant to say the number of subjects in that covariate class and prop was meant to say what fraction of the total analysis sample that constitutes. the remastered edition should get those numbers right.” The remastered edition prints the corrected column, computed from the maintained pipeline rather than typed, and each attribute’s classes sum to 1.00 across all 300 of them. The N column moves in 98 of the 787 rows for the same reason, in the four studies whose regressions carry more than a treatment indicator and an intercept. Nothing in the article’s text, tables or figures quotes a Prop value, so no other published number depends on it. The entry keeps its number, which is what the append-only rule protects; what changed is the remedy.

2. Table 1 reports each study’s sample size, not the number of observations the analyses use

The full list of studies, with the sample sizes used in the analyses reported here, is presented in Table 1.

Corrected:

The full list of studies, with **the number of respondents in each sample**, is presented in Table 1.

Table 1’s columns count everyone in the survey wave a study was fielded in. The treatment indicator is defined only for the respondents that study assigned, so the analyses use fewer: 69,599 observations against the 101,745 the two columns total. The reported sample is the larger of the two in 18 of the 27 pairs and exactly equal in 8. It is smaller in 1, Johnston and Ballard (2016), whose outcome is measured over several items per respondent, so its observations outnumber its subjects. The gap is widest at Murtagh et al. (2012), where Table 1 reports 5,243 and the analyses use 640, or 12 per cent.

Several studies were fielded in the same wave, and a shared wave is counted once for each of them: the 27 MTurk entries take only 14 distinct values, four studies reporting the same 1,913 between them. Summing the column therefore adds a wave in as many times as it carried experiments, which is how 101,745 arises from 28,872 distinct MTurk wave sizes and 37,003 original-study respondents. That total is the figure the abstract and the significance statement give for the individual survey responses the pairs encompass. It is a faithful sum of Table 1 and reproduces exactly; readers should know that it counts wave memberships rather than distinct people, and that a respondent counted in a study’s row was not necessarily assigned to that study’s experiment.

The article contradicts itself on this point rather than merely being imprecise. For Bergan (2012), Table 1 reports 1,206 in the original sample, while every attribute in the corresponding panel of appendix Table 1 sums to 396. The appendix is right. No published estimate

depends on the distinction: Table 1's counts are quoted in the article only as sample sizes, and each subgroup's own N is printed beside its estimate throughout the appendix.

16. The appendix's N column undercounts wherever a subgroup regression carries a third term

The N column gives the number of subjects in the associated covariate class.

Corrected:

The N column gives the number of subjects in the associated covariate class, **less one for each term in the regression beyond the treatment indicator and the intercept.**

The deposit builds the column as the regression's residual degrees of freedom plus two, which equals the number of observations only where the fit carries a treatment indicator and an intercept and nothing else. That holds for 689 of the 787 cells. In the remaining 98, spread over four study pairs, the regressions carry a further arm of the study's own factorial design, and each printed N falls short by between 1 and 3. In appendix Table 11, for instance, Hiscox (2006) prints 545 for age: 18 - 39 where 548 subjects are in that class.

The affected pairs are Hiscox (2006) (32 cells, appendix Table 11), McGinty, Webster and Barry (2013) (32 cells, appendix Table 16), Nicholson (2012) (20 cells, appendix Table 18), and Transue (2007) (14 cells, appendix Table 25). As with the Prop column above, the error is in the deposited code rather than in the typesetting, and the correction restates what the column contains rather than reprinting four tables to change one digit in each of 98 cells. No estimate, standard error, confidence interval or p-value is affected: the regressions themselves are fitted on every observation, and only the reported count is short.

The tables and figures

The five entries below correct something inside a float rather than a sentence: a column header, a study name, two legend titles and a rounded figure. Each was found by reading this paper's floats against the maintained pipeline's own output while a remastered edition of the article was being prepared. None of them touches an estimate, and every value plotted or tabulated beside them reproduces exactly.

5. Table 1's column header sets the platform's name differently from the rest of the paper

Published, p. 12442, Table 1:

Study	Original N	Mturk N	Slope (SE)	95% CI	N comparisons	Joint F
test	P value	Ref.				

Corrected:

Study	Original N	MTurk N	Slope (SE)	95% CI	N comparisons	Joint F	test P value	Ref.
-------	------------	---------	------------	--------	---------------	---------	--------------	------

The article sets the platform's name MTurk at every one of the ten places it appears in the running text, and the authors' own accepted manuscript sets the same header MTurk N. Only the published table lowers the T. The correction is to a heading and no estimate depends on it.

6. Table 1 names only the first of the fourth study's two authors

Published, p. 12442, Table 1:

Caprariello (2013)	825	2729	-4.38 (2.00)	[-7.83, -0.93]	16	0.63	27
--------------------	-----	------	--------------	----------------	----	------	----

Corrected:

Caprariello and Reis (2013)	825	2729	-4.38 (2.00)	[-7.83, -0.93]	16	0.63	27
-----------------------------	-----	------	--------------	----------------	----	------	----

Reference 27 is Caprariello PA, Reis HT (2013), and every other two-author study in Table 1 is named for both authors. The supplementary appendix's Table 4 and Fig. 2's own panel label both read Caprariello and Reis (2013); only Table 1's first column drops the second author. The estimates in the row are untouched and reproduce exactly.

7. Fig. 1's legend title sets CATEs in full capitals

Published, p. 12443, Fig. 1:

Difference in CATES

Corrected:

Difference in CATEs

The article writes CATEs, the plural of an initialism, everywhere else, including in this figure's own caption one line below the legend. The legend title is the plotting code's string and the capitals are a slip in it, not a different quantity.

8. Fig. 2's legend title carries the quotation marks of the string that set it, as well as the capitals

Published, p. 12444, Fig. 2:

‘Difference in CATES’

Corrected:

Difference in CATEs

The published legend reads with an opening single quotation mark at both ends, which is what a quoted R string prints when the quotes are inside the value rather than around it. Fig. 1 on the facing page has the same title without the marks. Neither the marks nor the capitals belong on the page.

15. Table 1 prints one slope to one decimal place where the other twenty-six are to two

Published, p. 12442, Table 1:

Hiscox (2006) | 1610 | 2972 | **2.5 (1.07)** | [0.94, 4.07] | 16 | 0.96 | 34

Corrected:

Hiscox (2006) | 1610 | 2972 | 2.50 (1.07) | [0.94, 4.07] | 16 | 0.96 | 34

The estimate is 2.5035, which rounds to 2.50 (1.07) and is printed correctly to the number of digits it carries; the deposited code rounds and then lets a trailing zero fall off, so this is the only cell of the column set to one decimal place. The value is unchanged and no reading of the table changes. It is listed because a remastered edition regenerates the table from the code and would otherwise print 2.50 (1.07) without saying why.

The reference list

The eight entries below correct the article's reference list. All were found by an audit that sends every printed entry whole to Crossref and checks the authoritative record back into it, and each is verified against the record named in the note. No cited work is missing and no citation in the text points at the wrong work: what is wrong is how the list prints an author's name, a year, a publisher's city, or a volume and page range.

Three entries were checked and left standing. Reference 33 (Gash and Murakami 2009) was flagged for a missing surname, which turned out to be Crossref matching a different work entirely. Reference 42 (Parmer 2011) prints a title the Time-Sharing Experiments record words differently, but no dissertation record could be retrieved to settle which the dissertation itself carries. Reference 13 (Coppock) prints no year at all, because at press the

article was forthcoming; supplying one would be updating the citation rather than correcting it.

3. Arthur Lupia's given name is printed as his surname in reference 1

Published, p. 12445:

1. Druckman JN, Green DP, Kuklinski JH, **Arthur L** (2006) The growth and development of experimental research in political science. *Am Polit Sci Rev* 100:627–635.

Corrected:

1. Druckman JN, Green DP, Kuklinski JH, Lupia A (2006) The growth and development of experimental research in political science. *Am Polit Sci Rev* 100:627–635.

Crossref, doi 10.1017/s0003055406062514: Druckman, Green, Kuklinski, Lupia. The fourth author's given name was promoted to a surname and his surname dropped, so the entry reads as though its fourth author were named Arthur.

4. The same author is printed the same way in reference 2

Published, p. 12445:

2. Druckman JN, Green DP, Kuklinski JH, **Arthur L** (2011) *Cambridge Handbook of Experimental Political Science* (Cambridge Univ Press, New York).

Corrected:

2. Druckman JN, Green DP, Kuklinski JH, Lupia A (2011) *Cambridge Handbook of Experimental Political Science* (Cambridge Univ Press, New York).

The four named are the handbook's editors. Crossref, doi 10.1017/cbo9780511921452.001, the opening chapter they wrote for it: Druckman, Green, Kuklinski, Lupia.

9. Reference 7 gives the year of the advance version and the volume and pages of the printed one

Published, p. 12445:

7. Aronow PM, Samii C (**2015**) Does regression produce representative estimates of causal effects? *Am J Polit Sci* 60:250–267.

Corrected:

7. Aronow PM, Samii C (2016) Does regression produce representative estimates of causal effects? *Am J Polit Sci* 60:250–267.

Crossref, doi 10.1111/ajps.12185: published online 28 April 2015 and in print in January 2016 as volume 60, pages 250 to 267. The entry's own volume and pages are the printed ones, so the year printed beside them should be the printed year.

10. Reference 14 prints a stray digit inside the year

Published, p. 12445:

14. Krupnikov Y, Levine AS (**1 2014**) Cross-sample comparisons and external validity. *J Exp Polit Sci* 1:59–80.

Corrected:

14. Krupnikov Y, Levine AS (2014) Cross-sample comparisons and external validity. *J Exp Polit Sci* 1:59–80.

Crossref, doi 10.1017/xps.2014.7: *Journal of Experimental Political Science* 1, pages 59 to 80, 2014. The 1 corresponds to no field the entry does not already carry; the issue number appears to have leaked into the year.

11. Reference 16 gives Deming's publisher the wrong city

Published, p. 12445:

16. Deming WE (1943) *Statistical Adjustment of Data* (Wiley, **Oxford, UK**).

Corrected:

16. Deming WE (1943) *Statistical Adjustment of Data* (Wiley, New York).

John Wiley and Sons published the 1943 first edition in New York, with Chapman and Hall handling London; the Open Library record for the 1943 printing gives New York and London and no Oxford. Wiley's Oxford imprint is decades later than the book.

12. Reference 23 gives the year of the advance version and the volume and pages of the printed one

Published, p. 12446:

23. Deaton A, Cartwright N (2017) Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine* 210:2–21.

Corrected:

23. Deaton A, Cartwright N (2018) Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine* 210:2–21.

Crossref, doi 10.1016/j.socscimed.2017.12.005: volume 210, pages 2 to 21, in the August 2018 issue. The 2017 in the doi suffix is the date the advance version was posted, and the entry's volume and pages are the printed ones.

13. Reference 38 gives the year of the advance version and the volume and pages of the printed one

Published, p. 12446:

38. Levendusky M, Malhotra N (2015) Does media coverage of partisan polarization affect political attitudes? *Polit Commun* 33:283–301.

Corrected:

38. Levendusky M, Malhotra N (2016) Does media coverage of partisan polarization affect political attitudes? *Polit Commun* 33:283–301.

Crossref, doi 10.1080/10584609.2015.1038455: published online 27 July 2015 and in print in April 2016 as *Political Communication* volume 33, pages 283 to 301.

14. Reference 47 prints the wrong volume and drops a digit from the first page

Published, p. 12446:

47. Thompson SC, Schlehofer MM (2014) Undermining optimistic denial reactions to domestic and campus emergency warning messages. *Appl Psychol Health Well Being* 2:19–213.

Corrected:

47. Thompson SC, Schlehofer MM (2014) Undermining optimistic denial reactions to domestic and campus emergency warning messages. *Appl Psychol Health Well Being* 6:192–213.

Crossref, doi 10.1111/aphw.12024: Applied Psychology: Health and Well-Being volume 6, issue 2, pages 192 to 213, 2014. The printed entry has the issue number in place of the volume and the first page short by a digit, so the range it prints spans nearly two hundred pages and points at nothing.