Errata

On Blair, Cooper, Coppock and Humphreys (2019), *American Political Science Review* 113(3)

The maintained replication code behind these errata is at https://github.com/active-maintenance-aec/blair_cooper_coppock_humphreys_2019 and is archived at Zenodo at https://doi.org/10.5281/zenodo.24168278, a DOI that always resolves to the latest version.

This note lists 57 corrections, and five items it flags without correcting, in "Declaring and Diagnosing Research Designs," published in *American Political Science Review* 113(3):838–859, doi 10.1017/S0003055419000194, and in its online appendix. Most are slips of wording, spelling, and typography, or details of the reference list. The rest correct a printed table, a figure caption, code printed in the article or appendix, a cross-reference, or a quantity stated in the text. Three entries correct a printed table (4, 11, and 56), and four flag claims in the article's main text (5, 13, 14, and 57).

The article's conclusions about discovery stand, and two other claims in its main text are not supported by the appendix's own tables. Entry 1 corrects both rows of the discovery table of appendix Section 3.3, which the deposited code screened on a t statistic where its text means the p-value. Under the corrected rows the statements on page 851 hold: principled discovery has lower bias and not necessarily lower error. Entry 11 flags the claim on page 848 that regression performs as well as or better than the Boolean algorithms in QCA, for which the appendix's table prints no support, and entry 15 flags the claim on page 855 that the interaction strategy is strictly dominated, which the appendix's table contradicts at a correlation of 0.8. Neither claim is repaired here, because a repair would rewrite the argument it supports. Entry 6 flags the rounding of two exaggeration ratios in the text, and entry 9 flags two percentages on page 846 that the design does not give.

Entries 52 and 53 are details of the reference list found by an audit that sends every printed reference entry whole to Crossref and checks the authoritative record back into it. The others were found by reading the article and its appendix word by word and by comparing the published text with its proofs, which alone produced entries 16, 26, 27, and 28. Each quotation was checked against an image of the published page, and each reference against the Crossref record named in its note. The corrected values in entries 1, 3, 4, 5, and 8 are read from the maintained replication's output, or computed from the design, at the time this document is rendered. The code supporting the maintained analysis is at github.com/active-maintenance-aec/blair_cooper_coppock_humphreys_2019.

1. The discovery code screens on the t statistic where it means the p-value, which moves both rows of the table

Published, app. pp. 38 to 40, and p. 851:

> train_pval = tidy(lm_robust(Y ~ Z * X, subset = train == 1))[4, **4**], ... all_pval = tidy(lm_robust(Y ~ Z * X))[4, **4**], ... Full sample Z:X **-0.25 0.32 0.29 0.05 1.00** ... Test sample Z:X -0.00 0.41 **0.32 0.05 0.78**

Corrected:

> train_pval = tidy(lm_robust(Y ~ Z * X, subset = train == 1))[4, **"p.value"**], ... all_pval = tidy(lm_robust(Y ~ Z * X))[4, **"p.value"**], ... Full sample Z:X **0.24 0.37 0.23 1.00 0.00** ... Test sample Z:X -0.00 0.41 **0.14 0.04 0.77**

The columns are Bias, RMSE, Frequency, False Pos, and False Neg. The code on p. 38 computes the two screening variables as `tidy(...)[4, 4]`. Column 4 of the `tidy()` output is `statistic`, not the p-value that the variable names `train_pval` and `all_pval`, the tests `< 0.05` on p. 39, and the prose on p. 40 describe. The archived versions of estimatr from 0.10.0 (July 2018) through 0.20.0 (September 2019) have `statistic` in that column, and the deposit dates from February 2019. The deposit's saved diagnosis, `rds_files/3_10_discovery_5000_sims.rds`, shows what the published run did. The 1,438 Full sample estimates it keeps are all at or below 0.015, with a median p-value of 0.55 and only 4.8 percent below .05, so the published run kept estimates by t, not by p. The Test sample row uses the same screen, so both rows are affected. The corrected code reads `[4, "p.value"]` in both lines.

The corrected rows are the maintained replication's: the same simulated draws, screened on the p-value, with 50,000 draws because the conditional bias moves in the second decimal at the published 5,000. Screening the same draws on the statistic, as printed, gives Full sample -0.25 0.32 0.28 0.05 1.00 and Test sample 0.00 0.41 0.32 0.05 0.76, within simulation error of the printed rows, which is why the printed rows are attributed to the screen. Both maintained scripts are in the repository: `maintained/table_a18_discovery.R` reproduces the printed table from the printed code, and `maintained/table_a18_discovery_intended.R` computes both screens on the same draws and writes the corrected values. Both use complete random assignment, as the appendix describes and as the deposit's `declare_assignment(prob = 0.5)` produced. The printed code is on appendix pp. 38 and 39 and the table on pp. 39 and 40. The standard errors the table prints are not corrected here: they belong to the printed values, and the maintained table computes none for the corrected ones.

The article's conclusions about discovery stand. Page 851's two statements, lower bias from principled discovery and mean squared error that is not necessarily lower, hold under the corrected rows: the Test sample's bias is -0.00 against 0.24 for the Full sample, and its RMSE is 0.41 against 0.37. Page 40's prose is right as written. Under the comparison method the false positive rate is 1.00 and the false negative rate is 0.00, where the printed row says 0.05 and 1.00 and so contradicts the prose. The principled method's false positive rate is 0.04, and it often fails to reject, with a false negative rate of 0.77. Its estimates are noisier: the spread

of the errors around the bias, the square root of the squared RMSE less the squared bias, is 0.41 against 0.28. It is less likely to produce a result at all, with a frequency of 0.14 against 0.23, where the printed table has the Test sample more frequent, 0.32 against 0.29. The Full sample's bias also changes sign, from negative to positive, which is the exaggeration from conditioning on significance that p. 40 describes.

2. The appendix gives the replication population as 50 clusters of 2,500 households where its code builds 300

Published, app. p. 41, Section 4.1:

> The data are hierarchical – there are 2500 households in each of **50** clusters. The resulting **125,000** row dataset is the population from which subjects will be sampled.

Corrected:

> The data are hierarchical – there are 2500 households in each of **300** clusters. The resulting **750,000** row dataset is the population from which subjects will be sampled.

The code below the sentence builds the catchment-area level with `add_level(N = 300, ...)` and the household level with `add_level(N = 2500, ...)`, which is 750,000 rows. The 50 is the number of clusters in the original study, `N_catchment_areas <- 50`, which the sampling step on p. 43 draws from the 300.

3. Three statements beneath the appendix's mortality table do not match the table

Published, app. p. 44, Section 4.5:

> we see that the true population average treatment effect is **0.006 percentage points**. In our simulations, we estimate the true standard error to be **0**, which is close to the standard error reported in the original paper of 0.026. ... only **0.1%** of simulations returned a statistically significant result.

Corrected:

> we see that the true population average treatment effect is **-0.06, or a reduction of 6 percentage points**. In our simulations, we estimate the true standard error to be **0.04**, which is close to the standard error reported in the original paper of 0.026. ... only **29%** of simulations returned a statistically significant result.

The table above the sentences prints a mean mortality estimand of -0.06 and a power of 0.29. The estimand is a difference in proportions (`((1 - mean(infant_alive_Z_1)) - (1 - mean(infant_alive_Z_0))`, p. 42), so -0.06 is a reduction of six percentage points, which 0.006 does not match. The table has no column for the standard error, so the second figure cannot be checked on the page. The maintained replication measures it as the standard deviation of the estimates across the simulations, 0.04, which is not very close to the original paper's 0.026, so the comparison the sentence draws is looser than it reads. The correction leaves that comparison to the authors. The maintained power is 0.30, within simulation error of the printed 0.29, and 0.1% matches neither.

4. The appendix says the unequal-probability design is biased in a table that prints a bias of 0.00

Published, p. 849, and app. p. 31:

> Interestingly, with the design in hand, it is easy to see that **unbiasedness is lost** in a design in which different values of xi are assigned with differing probabilities. ... However, as the unequal probability design shows, even with random assignment of Z to Y, **the linear estimator can be biased**.

The table on appendix p. 31 prints a bias of 0.00, with a standard error of 0.01, for every row, both the equal-probability design (`design_1`) and the unequal-probability design (`design_2`). The maintained replication gives -0.012 for `design_1`, where no bias is expected, and 0.008 for `design_2`, each with a standard error of about 0.006. Neither is distinguishable from zero in 5,000 simulations, and the equal-probability design lies farther from zero than the unequal one. Whether the estimator is biased under unequal probabilities is a property of the design that the printed table cannot show, and the two sentences present it as something the diagnosis shows.

5. The Bayesian posterior parameters are one too small, which moves the flat-prior row of the table

Published, app. pp. 10 and 11, Section 3.1.2:

> alpha <- n_successes + **alpha_0 - 1** ... beta <- n_trials - n_successes + **beta_0 - 1** ... 10 flat priors **0.46 0.14 0.00 -0.04 -0.15**

Corrected:

> alpha <- n_successes + **alpha_0** ... beta <- n_trials - n_successes + **beta_0** ... 10 flat priors **0.47 0.13 0.01 -0.03 -0.16**

The columns are Mean Est, Mean Sd, Bias, Mean Shift, and SD Shift. A beta prior with parameters alpha_0 and beta_0 updates to a beta posterior with parameters alpha_0 plus the successes and beta_0 plus the failures, so the printed code's `- 1` makes the posterior one count too small on each side and, for the flat prior, turns the posterior mean into the sample proportion. The corrected row is computed from the sampling distribution of the number of successes in 10 draws, with the design's mean success probability. The same computation with the `- 1` reproduces the printed flat row and the printed informative row to the printed digits, which confirms that it describes the table. The informative row, 0.49 0.09 0.03 -0.01 -0.02, does not change at two decimals. The table's n = 100 rows repeat its n = 10 rows in all five columns, which is a separate defect: the declaration writes the sample size as a literal inside `declare_sampling()`, where `redesign()` cannot reach it. Entry 8 corrects those rows.

6. The exaggeration ratios are rounded to "two" and "four" in the text

Published, p. 855, and app. p. 46:

> on average, statistically significant estimates tend to exaggerate the true effect of the intervention on mortality **by a factor of two** and on weight-for-age **by a factor of four**. ... limiting attention to statistically significant effects provides estimates that are **twice as large** in absolute value as the true effect size on average. ... the original study risked exaggerating the size of the weight effect by a factor of 4, and the mortality effect by a factor of almost 2.

The table on appendix p. 46 prints exaggeration ratios of 1.58 for mortality and 3.73 for weight-for-age. "A factor of two" and "twice as large" are read from 1.58, and "four" from 3.73. The appendix's own sentence says "almost 2" and "4", so the rounding looks deliberate. No correction is offered. A reader who wants the figures should take 1.6 and 3.7.

7. The critical value for the p-value is printed as 0.005

Published, p. 842:

> Equivalently, it is the probability that the p-value is lower than a critical value (e.g., **0.005**).

Corrected:

> Equivalently, it is the probability that the p-value is lower than a critical value (e.g., **0.05**).

8. The n = 100 rows of the Bayesian table repeat its n = 10 rows

Published, app. p. 11, Section 3.1.2.2:

> 100 flat priors 0.46 **0.14** 0.00 -0.04 **-0.15** 100 informative priors **0.49 0.09 0.03 -0.01 -0.02**

Corrected:

> 100 flat priors 0.46 **0.05** 0.00 -0.04 **-0.24** 100 informative priors **0.47 0.05 0.01 -0.03 -0.06**

The columns are Mean Est, Mean Sd, Bias, Mean Shift, and SD Shift. The printed rows repeat the n = 10 rows in all five columns. The code on appendix p. 10 writes the sampling step as `declare_sampling(n = 10)`, so `redesign(bayesian_design, n = c(10, 100))` on p. 11 has no sample size to change and both designs draw ten units. The corrected rows are read from the maintained replication, which names the sample size and keeps the printed posterior parameters, so this entry corrects this defect alone. The correction of entry 5 changes neither row at two decimals.

9. The Bayesian design does not give the 40% and 33% stated for it

Published, p. 846:

> In terms of the bias-variance tradeoff, the informative priors decrease the posterior standard deviation by **40%** relative to the uniform priors, but increase the bias by **33%**.

The direction of both comparisons holds, and the sizes do not. The design of appendix Section 3.1.2 gives a fall in the posterior standard deviation of 35% with 10 units and 8% with 100. The flat prior's bias in that design is zero, because the printed `- 1` of entry 5 makes its posterior mean the sample proportion, so no percentage increase over it exists. With the correction of entry 5, the fall is 33% and 8%, and the informative prior's bias is about 4.0 and 8.5 times the flat prior's. The note prints no replacement because the comparison behind 40% and 33% cannot be identified.

10. The Figure 1 caption swaps the line styles and markers of the two designs

Published, p. 850, Figure 1 caption:

> Bias (left), root mean-squared-error (center), and power (right) are displayed for two assignment strategies, a **2 × 2 treatment arm factorial design (black solid lines; circles) and a three-arm design (gray dashed lines; triangles)** according to varying interaction effect sizes specified in the potential outcomes function (x axis).

Corrected:

> Bias (left), root mean-squared-error (center), and power (right) are displayed for two assignment strategies, a **three-arm design (black solid lines; circles) and a 2 × 2 treatment arm factorial design (gray dashed lines; triangles and squares)** according to varying interaction effect sizes specified in the potential outcomes function (x axis).

The legend above the caption draws the three-arm design with a black solid line and circles. It draws the 2 × 2 factorial design with a gray dashed line, with triangles for the main-effect series and squares for the interaction-effect series. The caption gives each design the other's line style and markers, and names triangles alone for the factorial design.

11. The claim that regression performs as well as or better than the Boolean algorithms has no printed support

Published, p. 848, and app. pp. 21 and 22:

> Using regression to perform QCA **is equivalent to** QMC and CCubes when there is no error, and **even slightly outperforms** these algorithms (on the diagnosands we consider) in the presence of measurement error. ... In such cases, the saturated regression approach **slightly outperforms** the Boolean minimization algorithms.

The table on appendix p. 21 prints Correct 0.00 and Failure 1.00 for Least Squares in all four of its rows, which are error rates 0 and 0.2 at N of 3 and 50. The other two strategies print 0.10 and 0.32 for QMC and 0.10 and 0.05 for CCubes at N = 3 with no error, and 1.00 and 0.00 for both at N = 50 with no error. Page 22 has the second sentence and no table rows. A strategy that fails in every simulation cannot be read from the table as equivalent to the algorithms or as outperforming them, in either clause of the sentence on p. 848. The cause is in the code: the least squares strategy passes its predicted configurations to `QCA::simplify` inside a `try()`, `simplify` is not exported by the QCA package (checked under QCA 3.25), and every call fails, so every simulation returns "NO SOLUTION". The maintained replication reproduces the printed column for that reason, in four rows of `maintained/output/table_a11_qca_with_least_squares.csv`, all with a correct rate of 0 and a failure rate of 1. The note does not say what the least squares strategy would show if the call worked. The two claims have no support in what the article and appendix print.

12. The exaggeration ratio and the conditional estimator are described at the 0.05 level and computed at 0.1

Published, p. 855 (twice), and app. p. 47:

> given that the estimated effect is significant at the $\alpha = 0.05$ level. ... only if the CBO presence is significantly imbalanced among the 50 treatment and control clusters at the $\alpha = 0.05$ level; ... is "significantly" imbalanced among the 50 treatment and control clusters (at the $\alpha = .05$ level)

Corrected:

> given that the estimated effect is significant at the $\alpha = 0.1$ level. ... only if the CBO presence is significantly imbalanced among the 50 treatment and control clusters at the $\alpha = 0.1$ level; ... is "significantly" imbalanced among the 50 treatment and control clusters (at the $\alpha = .1$ level)

The code uses 0.1 in both places: the exaggeration ratio selects `p.value < 0.1` and the conditional estimator is declared as `function(data, strategy, alpha = 0.1)`, on appendix pp. 46 and 47. The sentence on appendix p. 46 introducing the diagnosis says "setting $\alpha = 0.1$". The printed symbol is α.

13. The appendix says respondents are more likely to vote than they say they will, where its code and the article say they overstate

Published, app. p. 8, Section 3.1.1, Model:

> People report their support truthfully, but are **more** likely to turn out to vote than they reveal.

Corrected:

> People report their support truthfully, but are **less** likely to turn out to vote than they reveal.

The code on the same page carries the comment "Subjects tend to overreport their probability of voting" and draws `likely_voter` from `pnorm(latent_voting + 0.75)`, and p. 846 of the article says the latent voters "overstate their true propensity to vote".

14. The nested analysis table does not show an effect of testing effort

Published, app. p. 28, Section 3.2.3:

> The table below **reveals a surprising feature of the nested analysis strategy: increasing the effort devoted to testing decreases our chances of happening on the right model. In fact, we are better off when more effort is devoted to qualitative theory building and less effort to qualitative theory testing.**

Corrected:

> The table below **does not show an effect of the effort devoted to testing or to building: its four rows are four simulations of the same design, and the proportions of simulations that arrive at the right model, 0.2780, 0.2858, 0.2782, and 0.2752, differ by no more than simulation error.**

The two effort parameters are used inside sampling rules that the design's handler calls, so `redesign()` cannot reach them, and the four rows are four runs of one design. The deposit's saved diagnosis has the same structure, with four rows that differ by no more than simulation error.

15. The claim that the interaction strategy is strictly dominated fails at a correlation of 0.8 in the appendix's own table

Published, p. 855, and app. p. 48:

> Including the interaction term is a **strictly dominated strategy** from the standpoint of reducing mean squared error: **irrespective of whether CBO presence is correlated with health outcomes or imbalanced**, the RMSE expected under this strategy is higher than under any other strategy. ... We note first that including the interaction term as done by the replicators is a strictly dominated strategy from the standpoint of reducing mean squared error: irrespective of whether CBO presence is correlated with health outcomes or imbalanced, the RMSE expected under this strategy is higher than under any other strategy.

The RMSE table on appendix pp. 48 and 49 gives RMSE, RMSE Balanced, and RMSE Imbalanced for each strategy. At a correlation of 0 the interaction strategy is the highest, 0.28, 0.28, and 0.27 against 0.27, 0.27, and 0.26 for each of the other three. At a correlation of 0.8 it is not: Ignore CBO prints 0.27, 0.23, and 0.49 and Interact CBO prints 0.24, 0.22, and 0.34, so interacting beats ignoring CBO in all three columns, and Include CBO is the lowest or tied for lowest throughout, at 0.19, 0.17, and 0.31. The article's later text agrees with the table, saying that when CBO presence is correlated with outcomes, strategies that include it "improve RMSE considerably". A correction would rewrite the argument that the sentence supports, the step to discounting the replicators' interaction result, so the sentence is flagged and left as printed.

16. The appendix describes independent assignment where its code assigns exactly 50 of 100 units

Published, p. 858, app., D(2):

> In this example **each sampled unit is assigned to treatment independently with probability 0.5**.

Corrected:

> In this example **exactly 50 of the 100 sampled units are assigned to treatment, by complete random assignment**.

The code beneath the sentence is `assignment <- declare_assignment(m = 50)`, and the sampling step above it is `declare_sampling(n = 100)`. Independent assignment with probability 0.5 would leave the number treated random.

17. The article's ten references to the Supplementary Materials use section numbers the appendix does not have

Published, pp. 846 to 851:

> Supplementary Materials Section **3.1**, we declare a model in which latent voters are likely to vote ... Supplementary Materials Section **3.2**, we declare a Bayesian descriptive inference design ... Supplementary Materials Section **3.3**. ... Supplementary Materials Section **3.4** we declare a QCA design ... Supplementary Materials Section **3.5**, we specify the estimands and analysis strategies ... Supplementary Materials Section **3.6**) ... Supplementary Materials Section **3.7**, we declare a design with a model in which three observable ... Supplementary Materials Section **3.8**, we declare and diagnose such a design ... Supplementary Materials Section **3.9**. ... Supplementary Materials Section **3.10**, we declare a design in which the model stipulates a treatment of interest

Corrected:

> Supplementary Materials Section **3.1.1**, we declare a model in which latent voters are likely to vote ... Supplementary Materials Section **3.1.2**, we declare a Bayesian descriptive inference design ... Supplementary Materials Section **3.2.1**. ... Supplementary Materials Section **3.2.2** we declare a QCA design ... Supplementary Materials Section **3.2.3**, we specify the estimands and analysis strategies ... Supplementary Materials Section **3.2.4**) ... Supplementary Materials Section **3.2.5**, we declare a design with a model in which three observable ... Supplementary Materials Section **3.2.6**, we declare and diagnose such a design ... Supplementary Materials Section **3.2.7**. ... Supplementary Materials Section **3.3**, we declare a design in which the model stipulates a treatment of interest

The article numbers the appendix's ten designs 3.1 to 3.10 in the order it discusses them. The appendix numbers them 3.1.1 (survey designs) and 3.1.2 (Bayesian descriptive inference) under 3.1, descriptive inference, then 3.2.1 (process tracing), 3.2.2 (qualitative comparative analysis), 3.2.3 (nested mixed methods), 3.2.4 (observational regression-based strategies), 3.2.5 (matching on observables), 3.2.6 (regression discontinuity), and 3.2.7 (experimental design) under 3.2, causal inference, and 3.3 (designs for discovery research). Each reference is matched to the design its sentence discusses. The article's 3.3 to 3.10 point at sections of the appendix that are other designs or do not exist.

18. The model's example names a disturbance that the equations do not define

Published, p. 840:

> (for example, $Y = bX + Z + \varepsilon Y$; $X = Z + \varepsilon X$, with Z, εX, $\boldsymbol{\varepsilon Z}$ distributed standard normal)

Corrected:

> (for example, $Y = bX + Z + \varepsilon Y$; $X = Z + \varepsilon X$, with Z, εX, $\boldsymbol{\varepsilon Y}$ distributed standard normal)

The equations define the disturbances εY and εX, and no εZ appears in them.

19. The inverse-probability-weighting estimator is called BFE-IPW once and IPW-BFE elsewhere

Published, p. 843:

> the third includes inverse-probability weighting to account for the heterogeneous assignment probabilities (**BFE-IPW**).

Corrected:

> the third includes inverse-probability weighting to account for the heterogeneous assignment probabilities (**IPW-BFE**).

Table 2 and the rest of the article print "IPW-BFE", for example "the IPW-BFE answer strategy" and "the IPW-BFE estimator".

20. "In additional we assume" should read "In addition, we assume"

Published, p. 840:

> In **additional** we assume functional relationships between them that allow for the possibility of confounding

Corrected:

> In **addition,** we assume functional relationships between them that allow for the possibility of confounding

21. "a researcher in correct" should read "a researcher is correct"

Published, p. 841:

> be interested in the probability that a researcher **in** correct in concluding that an intervention provides value for money.

Corrected:

> be interested in the probability that a researcher **is** correct in concluding that an intervention provides value for money.

22. The acknowledgments omit a comma between two names, and the first carries the wrong diacritic

Published, p. 838, acknowledgments:

> In addition, we thank Peter Aronow, Julian Brückner, Adrian **Duşa** Adam Glynn, Donald Green,

Corrected:

> In addition, we thank Peter Aronow, Julian Brückner, Adrian **Duș**a**,** Adam Glynn, Donald Green,

The corrected name has the comma-below s, ș, which the author asked for at proof with the note "please note the diacritical on the s". The published page prints a cedilla, ş, and no comma follows the name. The reference list prints the cedilla form in two entries, Duşa (2018) and Duşa and Thiem (2015). Those entries are not corrected here.

23. Footnote 6 says "no design could account for" where it means a design tool

Published, p. 843, note 6:

> 6 For example, no **design could** account for: the posited correlation between block size and potential outcomes; the sampling strategy; the exact randomization procedure; the formal definition of the estimand as the population average treatment effect; or the use of inverse-probability weighting.

Corrected:

> 6 For example, no **design tool could** account for: the posited correlation between block size and potential outcomes; the sampling strategy; the exact randomization procedure; the formal definition of the estimand as the population average treatment effect; or the use of inverse-probability weighting.

24. "the original replicators" should name the original authors and the replicators

Published, p. 855:

> In short, **the original replicators** make a set of contrasting claims about the true model of the world:

Corrected:

> In short, **the original authors and the replicators** make a set of contrasting claims about the true model of the world:

The sentence goes on to give the claims of "B&S", the original authors, and "D&M", the replicators, and the sentences before it call B&S "the original authors" and refer to Donato and Garcia Mosqueira as "the replicators".

25. A stray comma follows "Kreidler et al." in a parenthetical citation

Published, p. 843:

> (Green and MacLeod 2016; Groemping 2016; Guo et al. 2013; Kreidler et al., 2013)

Corrected:

> (Green and MacLeod 2016; Groemping 2016; Guo et al. 2013; Kreidler et al. 2013)

The other citations in the parenthetical have no comma between the author and the year.

26. A sentence on the article's last page is ungrammatical

Published, p. 859, app.:

> For example, to calculate the bias of the design **as a diagnosand is:**

Corrected:

> For example, to calculate the bias of the design**, we can declare a diagnosand as follows:**

The corrected wording is Jasper Cooper's, requested in the first proof round and not applied in the published text.

27. The printed code uses curly quotation marks and a minus sign that R rejects

Published, p. 845, Table 3, and p. 859, app. code:

> A <- declare_estimator(Y ~Z, estimand = **“ATE”**) ... e <- a_A[**“estimate”**] − a_M[**“estimand”**] ... model = lm_robust, estimand = **“ATE”**)

Corrected:

> A <- declare_estimator(Y ~Z, estimand = **"ATE"**) ... e <- a_A[**"estimate"**] - a_M[**"estimand"**] ... model = lm_robust, estimand = **"ATE"**)

R stops at the first curly quotation mark with `unexpected input`, so the code fails if typed as printed. The minus sign printed in the second line, and in the diagnosand formula `bias = mean(estimate - estimand)` on p. 859, is the Unicode minus, which R also rejects. The author asked for straight quotation marks in the final proof round, and the published tables and code print curly ones.

28. The research-note reference lowercases a word in a quoted title

Published, p. 856:

> Björkman Nyqvist, Martina, and Jakob Svensson. 2016. “Comments on Donato and Mosqueira’s (2016) ‘**additional Analyses**’ of Björkman and Svensson (2009).” Unpublished research note.

Corrected:

> Björkman Nyqvist, Martina, and Jakob Svensson. 2016. “Comments on Donato and Mosqueira’s (2016) ‘**Additional Analyses**’ of Björkman and Svensson (2009).” Unpublished research note.

The entry is an unpublished note, and a Crossref search finds no record for it, so the correction is to capitalization alone. The rest of the title is in title case.

29. The appendix spells “admissibility” as “admissability”

Published, app. pp. 1 and 6:

> 2.3 **Admissability** Criteria

Corrected:

> 2.3 **Admissibility** Criteria

The misspelling is in the contents on p. 1 and in the section heading on p. 6.

30. "it still calculable" is missing a verb

Published, app. p. 2:

> The estimand cannot be read from the dgp directly here but **it still** calculable.

Corrected:

> The estimand cannot be read from the dgp directly here but **it is still** calculable.

31. Two slips on the appendix's tool search page

Published, app. p. 4:

> **The data are heirarchical: individuals within blocks … all findings pertain to the tools we were able to locate at by that time point using the search methods described below.**

Corrected:

> **The data are hierarchical: individuals within blocks … all findings pertain to the tools we were able to locate by that time point using the search methods described below.**

32. "principle methods" should read "principal methods"

Published, app. p. 5:

> We used two **principle** methods to search for candidates.

Corrected:

> We used two **principal** methods to search for candidates.

33. "incporporate" should read "incorporate"

Published, app. p. 6:

> it could not **incporporate** heterogeneous block sizes or assignment probabilities

Corrected:

> it could not **incorporate** heterogeneous block sizes or assignment probabilities

34. "argmed" should read "armed"

Published, app. p. 13:

> it is more likely to observe the **argmed** group and the national army's takeover of natural resources when resources truly did cause the civil war

Corrected:

> it is more likely to observe the **armed** group and the national army's takeover of natural resources when resources truly did cause the civil war

35. "indictates" should read "indicates"

Published, app. p. 17:

> where lower case a indicates the absence of A and upper case B **indictates** the presence of B.

Corrected:

> where lower case a indicates the absence of A and upper case B **indicates** the presence of B.

36. "a an answer" should read "an answer"

Published, app. p. 19:

> The second is the probability that our answer strategy fails to provide **a an** answer at all (no solution to the minimization problem).

Corrected:

> The second is the probability that our answer strategy fails to provide **an** answer at all (no solution to the minimization problem).

37. "capable estimating" is missing "of"

Published, app. p. 21:

> the point that linear regression is **capable estimating** more parameters than it has degrees of freedom

Corrected:

> the point that linear regression is **capable of estimating** more parameters than it has degrees of freedom

38. "the the" is doubled

Published, app. p. 23:

> If **the the** regression is deemed satisfactory, we engage in small-N model-testing

Corrected:

> If **the** regression is deemed satisfactory, we engage in small-N model-testing

39. "relatonship" should read "relationship"

Published, app. p. 27:

> changes our probability of selecting the right theory of the X → Y **relatonship**.

Corrected:

> changes our probability of selecting the right theory of the X → Y **relationship**.

40. "arive" should read "arrive"

Published, app. p. 28:

> the proportion of simulations of the design in which we **arive** at the correct model.

Corrected:

> the proportion of simulations of the design in which we **arrive** at the correct model.

41. “are found are found” is doubled

Published, app. p. 38:

> Depending on whether significantly different subgroup are **found are found**, the difference in subgroup effects becomes a second question of interest, resulting from discovery.

Corrected:

> Depending on whether significantly different subgroup are **found**, the difference in subgroup effects becomes a second question of interest, resulting from discovery.

42. “it it” is doubled

Published, app. p. 40:

> The principled discovery method exhibits conventional rates for falsely rejecting a true null (0.05) though **it it** fails to reject the null quite often, due to weak power.

Corrected:

> The principled discovery method exhibits conventional rates for falsely rejecting a true null (0.05) though **it** fails to reject the null quite often, due to weak power.

43. “likey” should read “likely”

Published, app. p. 40:

> the principled strategy is somewhat less **likey** to produce a result at all

Corrected:

> the principled strategy is somewhat less **likely** to produce a result at all

44. "a mixed of types" should read "a mixture of types"

Published, app. p. 43:

> In the control group, the observed values are **a mixed** of types A and D, and in the treatment group, the values are a mixture of types B and D.

Corrected:

> In the control group, the observed values are **a mixture** of types A and D, and in the treatment group, the values are a mixture of types B and D.

45. "less-health infants" should read "less-healthy infants"

Published, app. p. 44:

> the assumption that **less-health** infants were the ones who are most likely to be of type B ("Beneficial")

Corrected:

> the assumption that **less-healthy** infants were the ones who are most likely to be of type B ("Beneficial")

46. Two misspellings in one sentence on the exaggeration ratio

Published, app. p. 45:

> We here examine the **exggeration** ratio, and in particular compare it to the exaggeration **ration** one might expect from the replication exercise conducted by Raffler, Posner, and Parkerson (2019).

Corrected:

> We here examine the **exaggeration** ratio, and in particular compare it to the exaggeration **ratio** one might expect from the replication exercise conducted by Raffler, Posner, and Parkerson (2019).

47. "Gelman and Carlin (2014) has highlighted" disagrees in number

Published, app. p. 45:

> **Gelman and Carlin (2014) has highlighted** how low power can generate bias (even in experiments)

Corrected:

> **Gelman and Carlin (2014) have highlighted** how low power can generate bias (even in experiments)

48. Stray backslashes print in the appendix's description of the mortality potential outcomes

Published, app. p. 42:

> The second potential outcome is the **weight_for_age** of surviving infants. ... In logits, this is moving from plogis(1.81) = **86\%** to plogis(1.81 + 0.5) = **91\%**.

Corrected:

> The second potential outcome is the **weight_for_age** of surviving infants. ... In logits, this is moving from plogis(1.81) = **86%** to plogis(1.81 + 0.5) = **91%**.

The same page prints "approximately 86%" without a backslash.

49. The appendix heading for Section 4 drops the diaeresis from an author's name

Published, app. pp. 1 and 41:

> 4 **Bjorkman** and Svensson (2009) Design Replication

Corrected:

> 4 **Björkman** and Svensson (2009) Design Replication

The heading prints "Bjorkman" in the contents on p. 1 and above the text on p. 41, and the sentence below it on p. 41 prints "Björkman and Svensson (2009)".

50. "right-hand side" is printed as one word

Published, app. p. 17:

> The **righthandside** of the regression includes indicators for membership in all feasible configurations

Corrected:

> The **right-hand side** of the regression includes indicators for membership in all feasible configurations

51. Two more formulas in the printed code use a minus sign that R rejects

Published, p. 845, Table 3, and pp. 858 and 859, app. code:

> ATE = mean(Y_Z_1 − Y_Z_0) ... bias = mean(estimate − estimand)

Corrected:

> ATE = mean(Y_Z_1 **-** Y_Z_0) ... bias = mean(estimate **-** estimand)

Table 3 prints both formulas, p. 858 prints the first again, across a line break, and p. 859 prints the second again. Entry 27 corrects the minus between `a_A` and `a_M` in the same table for the same reason, and its note mentions the second formula on p. 859 without correcting it.

The reference list

52. The Green and Lin reference ends its page range one page early

Published, p. 857:

> PS: Political Science and Politics 49 (3): **495–9**.

Corrected:

> PS: Political Science and Politics 49 (3): 495–500.

Crossref, doi 10.1017/S1049096516000810, gives the article's pages as 495 to 500.

53. The Nosek et al. reference prints a first page in place of a page range

Published, p. 857:

> Science 348 (6242): **1422**.

Corrected:

> Science 348 (6242): 1422–5.

Crossref, doi 10.1126/science.aab2374, gives the article's pages as 1422 to 1425. The entries around it on the page print full ranges.

54. The Rennie reference prints the journal's name twice

Published, p. 857:

> Rennie, Drummond. 2004. "Trial Registration." **Journal of the American Medical Association**: The Journal of the American Medical Association 292 (11): 1359–62.

Corrected:

> Rennie, Drummond. 2004. "Trial Registration." **JAMA**: The Journal of the American Medical Association 292 (11): 1359–62.

Crossref, doi 10.1001/jama.292.11.1359, gives the journal as JAMA, volume 292, issue 11, with first page 1359. The journal's title is JAMA: The Journal of the American Medical Association, and the entry spells the first half out a second time.

55. The Gelman and Carlin reference runs the title and subtitle together

Published, p. 857:

> Gelman, Andrew, and John Carlin. 2014. "Beyond **Power Calculations Assessing** Type S (Sign) and Type M (Magnitude) Errors."

Corrected:

> Gelman, Andrew, and John Carlin. 2014. "Beyond **Power Calculations: Assessing** Type S (Sign) and Type M (Magnitude) Errors."

Crossref, doi 10.1177/1745691614551642, gives the title as "Beyond Power Calculations" and the subtitle as "Assessing Type S (Sign) and Type M (Magnitude) Errors", in Perspectives on Psychological Science 9 (6).

56. The Dawid reference lacks a space after an initial

Published, p. 857:

> Dawid, **A.Philip**. 2000. "Causal Inference without Counterfactuals."

Corrected:

> Dawid, **A. Philip**. 2000. "Causal Inference without Counterfactuals."

Crossref, doi 10.1080/01621459.2000.10474210, gives the author as A. P. Dawid, in the Journal of the American Statistical Association 95 (450): 407–24. The correction restores the space and keeps the name the entry gives.

57. The reference for the eQMC article splits the name eQMC

Published, p. 857:

> Duşa, Adrian, and Alrik Thiem. 2015. "Enhancing the Minimization of Boolean and Multivalue Output Functions with **e QMC**."

Corrected:

> Duşa, Adrian, and Alrik Thiem. 2015. "Enhancing the Minimization of Boolean and Multivalue Output Functions with **eQMC**."

Crossref, doi 10.1080/0022250x.2014.897949, gives the title as ending "With eQMC".

58. The Kreidler et al. reference lacks the colon before its page range

Published, p. 857:

> Kreidler, Sarah M., Keith E. Muller, Gary K. Grunwald, Brandy M. Ringham, Zacchary T. Coker-Dukowitz, Uttara R. Sakhadeo, Anna E. Barón, and Deborah H. Glueck. 2013. "GLIMMPSE: Online Power Computation for Linear Models with and without a Baseline Covariate." Journal of Statistical Software 54 (**10**) **1–26**.

Corrected:

> Kreidler, Sarah M., Keith E. Muller, Gary K. Grunwald, Brandy M. Ringham, Zacchary T. Coker-Dukowitz, Uttara R. Sakhadeo, Anna E. Barón, and Deborah H. Glueck. 2013. "GLIMMPSE: Online Power Computation for Linear Models with and without a Baseline Covariate." Journal of Statistical Software 54 (**10**)**: 1–26**.

Crossref, doi 10.18637/jss.v054.i10, gives volume 54 and issue 10 and carries no page range. The other journal entries in the list separate the issue from the pages with a colon.

59. The Morris, White, and Crowther reference prints no volume or pages

Published, p. 857:

> Morris, Tim P., Ian R. White, and Michael J. Crowther. 2019. "Using Simulation Studies to Evaluate Statistical Methods." Statistics in Medicine**.**

Corrected:

> Morris, Tim P., Ian R. White, and Michael J. Crowther. 2019. "Using Simulation Studies to Evaluate Statistical Methods." Statistics in Medicine **38 (11): 2074–102.**

Crossref, doi 10.1002/sim.8086, gives volume 38, issue 11, and pages 2074 to 2102. The issue is dated 20 May 2019, and this article was published online on 31 May 2019 (Crossref, doi 10.1017/S0003055419000194), so the full citation existed when the article appeared.

60. The appendix's reference for the eQMC article prints the name as "E Qmc"

Published, app. p. 49:

> Duşa, Adrian, and Alrik Thiem. 2015. "Enhancing the Minimization of Boolean and Multivalue Output Functions with **E Qmc**."

Corrected:

> Duşa, Adrian, and Alrik Thiem. 2015. "Enhancing the Minimization of Boolean and Multivalue Output Functions with **eQMC**."

Crossref, doi 10.1080/0022250x.2014.897949, gives the title as ending "With eQMC". Entry 57 corrects the same title in the article's reference list.

61. The Sekhon and Titiunik reference drops the accent in "Rocío"

Published, p. 858:

> Sekhon, Jasjeet S., and **Rocıo** Titiunik. 2016. "Understanding Regression Discontinuity Designs as Observational Studies."

Corrected:

> Sekhon, Jasjeet S., and **Rocío** Titiunik. 2016. "Understanding Regression Discontinuity Designs as Observational Studies."

The page prints a dotless i with no accent.

62. The appendix's Gelman and Carlin reference runs the title and subtitle together

Published, app. p. 49:

> Gelman, Andrew, and John Carlin. 2014. "Beyond **Power Calculations Assessing** Type S (Sign) and Type M (Magnitude) Errors."

Corrected:

> Gelman, Andrew, and John Carlin. 2014. "Beyond **Power Calculations: Assessing** Type S (Sign) and Type M (Magnitude) Errors."

Entry 55 corrects the same title in the article's reference list.