These files replicate the results in the final (January 2019) version of "Identification of and Correction for Publication Bias."

The raw data underlying the applications in the paper is contained in the "Applications" folder, with a separate folder for each application. These folders also contain cleaned and standardized versions of each dataset, where the data cleaning is carried out by the files in the "Data Cleaning" folder. Code for the numerical illustrations reported in the paper is contained in the "Illustrations" folder.

The results reported in the paper are replicated by the file "AllApplications.m." in the "Matlab" folder. Each specification reported in the paper and supplement corresponds to a choice of the variable "application," with the correspondence listed in comments in AllApplications.m. For this code to work, it is important that the "Matlab" folder be in a the same folder as the "Applications" and "FiguresandTables" folders.

Contents of Zip Folder: We next list the files and folders included in the replication zip file:

1. Applications (Folder): contains raw and cleaned data for empirical applications reported in the paper

   (a) Deworming (Folder): contains raw and cleaned data for deworming application based on Croke et al. (2016)

      i. Table1Export.xlsx: Raw data for application

      ii. DewormingTable1.csv: Extracts estimates and standard errors for deworming studies

      iii. DewormingLables.csv: Names of studies corresponding to estimates in DewormingTable1.csv

      iv. cleaned_deworming_data.csv: Deworming data cleaned by clean_deworming_data.m

      v. DewormingLables.csv: Names of studies sorted by clean_deworming_data.m

   (b) EconExperiments (Folder): contains raw and cleaned data for experimental economics application based on Camerer et al. (2016)

      i. ReplicationEconComplete.csv: Raw data for application, constructed as described in DataDetails.text

      ii. DataDetails.text: Describes the data in ReplicationEconComplete.csv

      iii. EconLabels.csv: Names of studies corresponding to estimates in ReplicationEconComplete.csv

      iv. cleaned_econ_data.csv: Econ lab experiment data cleaned by clean_econ_replication_data.m

      v. sorted_names.csv: Names of studies sorted by clean_econ_replication_data.m

   (c) MinimumWagev2 (Folder): Contains raw and cleaned data for minimum wage application based on Wolfson and Belman (2015)

      i. Belman-Wolfson Minimum Wage Encoded.xlsx: Raw data for application

      ii. Data Details.text: A description of Belman-Wolfson Minimum Wage Encoded.xlsx

      iii. BWWP-MinWageData.csv: numerical raw data for application

      iv. BWWP-MinWageText.csv: text data for application

      v. cleaned_minwage_data.csv: Minimum wage data cleaned by clean_min_wage_data.m

     vi. sorted_text.csv: Text data sorted by clean_min_wage_data.m

(d) NatureScienceExperiments (Folder): contains raw and cleaned data for social science experiments application based on Camerer et al. (2018)

      i. D3 - ReplicationResults.csv: Raw data for application

     ii. SelectedNatureScienceColumns.csv: Columns of numerical data from D3 - ReplicationResults.csv that used in analysis

    iii. NatureScienceLabels.csv: Text data from D3 - ReplicationResults.csv that used in analysis

    iv. cleaned_naturescience_data.csv: Social science experimental data cleaned by clean_naturescience_replication_data.m

     v. sorted_names.csv: Author names for social science experiments sorted by clean_naturescience_replication_data.m

    vi. sorted_1stAuthor.csv: Names of first authors for social science experiments sorted by clean_naturescience_replication_data.m

(e) PsychExperiments (Folder): contains raw and cleaned data for psychology experiments application based on Open Science Collaboration (2015)

      i. rpp_data.csv: Raw data for application

     ii. rpp_data_codebook.csv: data description for rpp_data.csv (from from Open Science Collaboration (2015))

    iii. SelectedPsychColumns.csv: Numerical data from rpp_data.csv that used in analysis

    iv. PsychLabels.csv: text data from rpp_data.csv that used in analysis

     v. cleaned_psych_data.csv: Psychology experimental data cleaned by clean_psych_replication_data.m

    vi. sorted_names.csv: Author names for psychology experiments sorted by clean_naturescience_replication_data.m

    vii. sorted_1stAuthor.csv: Names of first authors for psychology experiments sorted by clean_naturescience_replication_data.m

2. Data Cleaning (Folder): contains Matlab code for cleaning data

(a) clean_deworming_data.m: Cleans data for deworming application

(b) clean_econ_replication_data.m: Cleans data for econ lab experiments application

(c) clean_min_wage_data.m: Cleans data for minimum wage application

(d) clean_naturescience_replication_data.m: Cleans data for social science experiment application

(e) clean_psych_replication_data.m: Cleans data for psychology experiments application

3. FiguresandTables (Folder): empty folder to hold results produced by running AllApplications.m (but needs to be present for AllApplications.m to run)

4. Matlab (Folder): contains code for replicating results reported in the paper

    (a) AllApplications.m: Replicates the applications reported in the paper. The remaining .m files in this folder are used by AllApplications.m

    (b) Clustered_covariance_estimate.m: implements a cluster-robust variance estimator

    (c) DescriptiveStats.m: Produces scatterplots and histograms based on raw data

    (d) DescriptiveStatsCombined.m: Produces figures containing both scatterplots and histograms based on raw data

    (e) EstimatingSelection.m: Implements estimation for specifications considered in paper

    (f) HorizontalBars.m: Generates corrected inference plots based on estimated selection model

    (g) MetaRegressionTable.m: Formats tables of meta-regression coefficients

    (h) MetastudyGMMObjective.m: Calculates GMM objective function in metastudy applications

    (i) MetastudyMoments.m: Calculates GMM moment conditions in metastudy applications

    (j) ReplicationAnalyticLogLikelihoodControls.m: Calculates log likelihood in replication applications with controls

    (k) ReplicationGMMObjective.m: Calculates GMM objective function in replication applications

    (l) ReplicationMoments.m: Calculates GMM moment conditions in replication applications

    (m) ReplicationpzthetaLogLikelihood.m: calculates likelihood in models where publication probability may depend on both the results and the true parameter value

    (n) RobustVariance.m: Calculates misspecification-robust variance estimator

    (o) SelectionTable.m: Formats tables of estimated coefficients for selection model

    (p) SelectionTableGMM.m: Formats tables of estimated coefficients for selection model estimated by GMM

    (q) Step_function_normal_cdf.m: Calculates distribution function for normal model with a step function publication probability

    (r) VariationVarianceLogLikelihoodControls.m: Calculates log likelihood in metastudy applications with controls

    (s) VariationVarianceLogLikelihoodtdist_logit: Calculates log likelihood in metastudy application with time trends (for minimum wage application)

    (t) VariationVarianceLogLikelihoodtdist: Calculates log likelihood in metastudy application with t distribution for true effects (for minimum wage application)

    (u) Illustration (Folder): Calculates the numerical illustrations in the paper

        i. IllustratingIdentification.m: Produces the figures used to illustrate the replication and metastudy identification strategies in the paper

ii. Normal Interval Truncation (Folder): Produces the figures used to illustrate the effect of publication bias, and corrections, in running example

    A. Publication_bias_plot.m: Simulates bias and coverage distortion from selective publication in running example

    B. Combined_truncated_inference_plot.m: produces plots showing the effect of correcting Bayesian and frequentist inference for selective publication

    C. Interval_truncation_normal_cdf.m: Calculates normal cdf when the publication probability takes form of toy example

    D. Interval_truncation_normal_pdf.m: Calculates normal pdf when the publication probability takes form of toy example

iii. Diff in Diffs (Folder): Contains files for producing the results for difference-in-differences models reported in the supplement

    A. Diff_in_Diff_Bias_plot.m: Simulates bias and coverage distortion from selective publication in difference-in-differences example

    B. truncation_Estimator_plot_DiffinDiff.m: Calculates and plots corrected estimator in difference-in-differences example

    C. Step_function_multivariate_normal_cdf.m: Calculates multivariate truncated normal cdf in difference-in-differences example

    D. product_partition_calculate.m: Calculates publication probabilities in difference-in-differences setting