Re-expressing coefficients from regression models for inclusion in a meta-analysis

Matthew W. Linakis,[a] Cynthia Van Landingham,[b] Matthew P. Longnecker[a]

[a] Ramboll U.S. Consulting, Raleigh NC 27612

[b] Ramboll U.S. Consulting, Monroe LA 71201

Corresponding Author: Matthew P. Longnecker, Ramboll U.S. Consulting. 3214 Charles B Root Wynd #130, Raleigh, NC, 27612.

Ph: 919-765-8029

Em: mlongnecker@ramboll.com

Keywords: Meta-analysis as Topic, re-expression, log-transformation, Systematic Review, Regression Coefficients

Abstract

Background: Quantitative summaries of study results, or meta-analyses, have proven useful in several fields, but pose special challenges when applied to non-randomized studies of exposure in relation to outcome. One such issue arises when summarizing regression results when some studies have log-transformed exposure while others have not. Methods of re-expression have recently been developed, with the goal of generating comparable results across studies regardless of data transformation. We examined the validity of three re-expression methods using simulations and real data examples.

Methods: Simulations were developed to examine the validity of the re-expression methods when the exposure had a skewed distribution. We also identified 15 published studies for which we were able to either access the data and replicate the result, or which had provided regression coefficients (β) from analyses with both log-transformed and untransformed exposure data. For the simulated and real data we re-expressed β as though it had been based on untransformed data, and vice versa. We compared the re-expressed results to fitted regression coefficients. The proportional difference from the fitted β was used to quantitate bias. We also examined the effect of influential observations on the success of the re-expression using the real data examples.

Results: In the simulated and real data examples, all three re-expression methods usually gave biased results. The method of Rodriguez-Barranco et al. (2017) tended to give less biased results but with a wider range of bias. For example, in the real data, when re-expressing results as if they had been fit to untransformed exposure, the median estimate was 22% too high, with a range of -40% to 1,904%. In some analyses, the presence of influential observations had a large effect on the validity of the re-expression.

Conclusions: When the distribution of exposure is skewed, the re-expression methods examined are likely to give biased results. The bias varied by method, the direction of the re-expression, skewness,

influential observations, and in some cases, the median exposure. Meta-analysts using any of these re-expression methods may want to consider the uncertainty, the likely direction and degree of bias, and conduct sensitivity analyses on the re-expressed results.

Introduction

The results of a group of studies deemed comparable can be synthesized quantitatively using meta-analysis. The goals of meta-analysis are: "1) improve precision, … 2) answer questions not posed by individual studies, … and 3) settle controversies arising from apparently conflicting studies or to generate new hypotheses" [1]. To base the meta-analysis on all available data, "Results extracted from study reports may need to be converted to a consistent, or usable, format for analysis" [2]. Methods of converting data presented by authors into a suitable format for meta-analysis have been well developed for intervention studies, but for studies of exposure effects the available tools are somewhat limited.

A challenge that frequently arises in the meta-analysis of non-randomized studies of exposure effects is that the authors of the studies being synthesized have made different choices about how to represent a continuous-scale exposure in their analysis. For example, the exposure might be left in its original untransformed units, log-transformed, or categorized.

Various approaches to the problem of inconsistently-expressed effect estimates have been recommended or used in practice. Obtaining the raw data or asking authors to re-analyze their data are the ideal solutions, though not always practicable. When these options are not feasible, the results of studies that used transformed and original units can be analyzed separately [3,4], or the results from studies using the less-frequent approach are simply excluded from the meta-analysis [5]; neither of these approaches is optimal. Here we consider methods of re-expressing regression coefficients from linear models fit to the log-transformed exposure variable as the coefficient that would have been obtained had the authors left the exposure in its original units. We refer to this process as re-expression of β to an untransformed basis. We also consider the converse, i.e., how to re-express regression coefficients from linear models fit to the exposure variable in its original units as the coefficient that

would have been obtained had the authors log-transformed the variable before fitting the regression. Similarly, we refer to this process as re-expression of β to log basis.

An algebraic method of re-expressing regression coefficients was recently proposed and evaluated using simulated data by Rodriguez-Barranco et al. [6]; they found that in the setting of a log-transformed lognormally distributed independent variable that was re-expressed as what the authors would have gotten had the model been fit to the untransformed data, the re-expressed coefficient could be half the size of the true (fitted) coefficient. They recommended caution in applying their method when the distribution of the independent variable was markedly asymmetric. More recently, empiric methods of re-expressing coefficients from models fit to a log-transformed independent variable to approximate what would have been obtained if fit to an original unit continuous independent exposure have been presented [7,8]. Steenland et al. (2018) described their method as

> … iteratively minimizing the squared deviation of a new linear curve from the original logarithmic one, over a scale of 0 to 10 ng/ml PFOA [perfluorooctanoic acid], typical of studies in the general population. We also minimized squared deviation of a linear upper and lower confidence limit from the original logarithmic confidence interval curves. For any given study, the iteration was conducted by minimizing the sum of squares of the difference between the candidate linear curve and the logarithmic curve reported in the literature, across 10 points, at 1, 2…. through 10 ng/ml. Iteration began with an educated guess for a candidate linear curve that would approximate the logarithmic curves and proceeded by varying the candidate linear curve until the sum of squares of the differences were minimized.

Steenland et al. did not evaluate the validity of their method. Dzierlenga et al. (2020) used the same basic principle, though modified the method so that it was more flexible with respect to the original scale of the exposure variable. The modification used an algorithmic optimization over 6 points from the 25th to the 75th percentiles (25th, 35th…75th) of the estimated exposure distribution. Dzierlenga et al. evaluated the validity of their re-expression and found that it performed well. Their evaluation, however, included only 5 studies and one exposure (perfluorooctane sulfonic acid, PFOS). In addition to the above, we developed a third ("alternative") estimator that is algebraic but different than that of Rodriguez-Barranco et al., and introduce it below, in the methods section.

The goals of the present project were to evaluate the validity of re-expression of regression coefficients to an untransformed basis, and the converse. The validity was assessed for the Rodriquez-Barrranco et al. (RB) estimator, the Dzielenga et al. (Dz) estimator, and the alternative (Alt) estimator using simulated and real data examples.

Methods

<u>The estimators</u> Rodriquez-Barranco et al. (2017) presented two algebraic methods of re-expression of interest here, one for re-expression of β to an untransformed exposure, and one for re-expression of β to log basis [6]. They also presented related equations for re-expression that addressed the setting where the outcome variable was log-transformed. Their results suggested that transformation of the outcome was a much smaller source of bias in re-expression than was transformation of exposure, so we have focused on the latter as it seems to be a more pressing matter in meta-analyses. The equations below correspond to their Model B for re-expression of β to an untransformed x (Eq. 1) and Model A for re-expression of β to logged x (Eq. 2)(See Table 1 of their publication):

$$\beta_R = log_b\left(1 + \frac{c}{E[X]}\right) * \beta_{from\ model\ with\ \log(x)} \#(1)$$

$$\beta_R = (k-1) * E[X] * \beta_{from\ model\ with\ untransformed\ x} \#(2)$$

In equations 1 and 2, β_R is re-expressed beta, c is the absolute change in x (c = 1 unit of exposure in the present study), E[X] is the mean of the exposure, and k is the relative change in x (defined as the log base, and therefore equal to b). The algebraic method of re-expression described by Rodriguez-Barranco et al. (2017) was originally formulated to use the mean of the exposure distribution [6]. However, when exposure distributions are skewed, often medians rather than means are presented by original authors. With this in mind, we modified the procedure by substituting median values for means when applying the RB estimators.

The Dzierlenga et al. (2020) algorithm ("Dz estimator") for re-expression [8] was described in the introduction above. We note that this method can handle multiple different measures of exposure central tendency and spread such that when medians and quartiles were not available, but geometric or arithmetic means and standard deviations were, the algorithm will accept those as inputs. Furthermore, we have modified it for re-expression of β to log basis.

The alternative method of re-expression ("Alt estimator") that we developed was based on the principle of calculating, on the untransformed scale of exposure, the increment that represented a doubling, a 2.718-fold increase, or a 10-fold increase (i.e., one log unit, depending on the base of the log, *a*). This was done by subtracting or adding 0.5 units on the log scale to the log(median), back-transforming the results, and taking the difference (see Equation 3).

$$I = a^{log_a(median)+0.5} - a^{log_a(median)-0.5} \#(3)$$

In equation 3, *I* is the increment used to re-express from log to linear ($\beta_{Reported}/I$) or linear to log ($\beta_{Reported} * I$), *a* is the logarithm base, and median is the median of the exposure variable. R scripts/functions and

data files for applying each of these three re-expression methods are available in the supplemental materials (R_Script_for_Data_Analysis_and_Table_Generation.html).

The simulation we performed was an idealized scenario where a known function generated the values of the dependent variable y based on values of the independent variable x drawn from a $\log_e$ normal distribution. For simplicity, no error was added to y. This gave results that focused on the mathematics underlying each re-expression method. The scenario first examined was with the function $y = \beta_{log} \cdot \log_e(x)$. In the baseline scenario β_{log} was given a value of 1 and the variable x was selected at random 10^6 times from a $\log_e$ normal distribution with μ_{ln} = ln(median of x) and $\sigma_{log} = 0.5$. In each scenario a series of median values of x was selected so that it approximately spanned the range of medians reported in the real data examples (0.25, 0.5, 1, 2, 4, 8, 16, 32), described below. (In the real data examples, we fit a $\log_e$ normal distribution to the median and quartiles and estimated the σ; the median σ was 0.59 and the 95th percentile of σ was 0.98.) The linear model $y = \alpha + \beta_{fitted} \cdot x + \varepsilon$ was then fit to the generated data and then β_{log} was used, along with the median of x (and for the Dzierlenga method, the interquartile range of x), to estimate β_{fitted} using each of the re-expression methods, as described in the previous section.

After the first scenario baseline runs, we varied the following parameters one at a time, $\sigma_{ln} = 1$, $\beta_{log} = 0.5, 10$; we also examined the effect of using $\log_2(x)$ when generating y.

In the second scenario, the above set-up and procedures were the same except the function generating values of y was $y = \beta_{linear} \cdot x$, the model $y = \alpha + \beta_{fitted} \cdot \log(x) + \varepsilon$ was fitted to the data (or fitting the same model using $\log_2(x)$), and β_{linear} was used to estimate β_{fitted} using each of the re-expression methods.

To evaluate the validity of re-expression methods, we sought examples for various types of outcome (dichotomous, log-continuous, untransformed continuous), and a variety of environmental

agents as measured with an exposure biomarker. Environmental exposures measured with a biomarker frequently have skewed distributions with a long tail to the right. We first identified a series of published analyses based on data that were publicly available. Second, we identified a similar series of published analyses that did not have raw data available but that presented regression results obtained with and without log transformation of the exposure.

For the example data that involved our re-analysis of published results, we chose results that could efficiently be replicated to a reasonable degree of accuracy using the originally described methods. When the authors presented results for more than one outcome or more than one exposure in a report, in general we arbitrarily chose one result that was statistically significant for inclusion in the evaluation of validity; the exception was data for Xu et al. (2020), where two different outcomes, one continuous, and one dichotomous, were examined in relation to the same exposure, and the regression coefficients were statistically significant for both [9].

Using the simulated and example data, we applied the re-expression methods. We expressed the results as the proportional difference (pΔ) from the fitted coefficient, or $\beta_R/\beta_F - 1$, where β_R is the β estimated using the re-expression method and β_F is the fitted regression coefficient.

In the two example datasets where the proportional difference from the fitted β was largest, we explored whether the exclusion of influential observations affected the accuracy of the re-expression using the Dz estimator. In two additional examples datasets where the proportional difference from the fitted β was typical of other studies, we also examined the effect of excluding influential points on the validity of the re-expression with the Dz estimator. Influential observations were identified with a difference in beta analysis (change in beta with each observation excluded one at a time) performed on the regression using untransformed exposure. A t-test-like statistic was used to identify the 5% of points that were unusually influential ($|DFBETAS| > 2/\sqrt{n}$) [10].

Results

Simulations

A simulation with data generated with $y = \beta_{log} \cdot \log_e(x)$ and "baseline" parameters ($\beta_{log} = 1$ and $\sigma = 0.5$) is depicted in Figure 1 for a median x of 1 ($\mu = 0$). In this scenario the RB estimator slightly undershot the slope estimated from the fitted regression line, whereas the Alt and Dz estimators overshot the fitted slope, by a slightly greater magnitude.

When y was a function of $\log_e(x)$ in the baseline scenario, the fitted β from the linear model decreased with the median of x (Table 1, Panel A), and the model fit the data fairly well (R^2=0.88). The RB estimator underestimated the observed β when the median of x was ≤ 1, did well when the median of x was 1 or 2, and overestimated it with greater median x. The Alt estimator had a modest positive bias that was fairly constant. The Dz estimator behaved much like the Alt estimator.

With a more skewed distribution of x, the estimators substantially overshot the observed β. The RB estimator was less biased when median x was low (Table 1, Panel B). With sigma back to 0.5 and y generated with $\log_2(x)$, the results were like those in the baseline simulation (Table 1, Panel C). When using $\log_e(x)$ to generate y, as in the baseline scenario, but with the β_{log} set to 0.5, all three estimators behaved as in the baseline scenario (Table 1, Panel D). With β_{log} set to 10, all three estimators behaved as in the baseline scenario (Table 1, Panel E).

For the simulations where $y = \beta_{linear} \cdot x$ and baseline parameters ($\beta_{linear} = 1$, $\sigma = 0.5$), the RB estimator consistently overshot the β fitted to $\log_e(x)$, whereas the Alt and Dz estimators were consistently too low, but by only a small amount (Table 2, Panel A). When the skewness of the distribution of x was increased ($\sigma = 1$), the RB estimator performed well whereas the Alt and Dz estimators were uniformly too low by 35%-37% (Table 2, Panel B). With $\sigma = 0.5$ and data analyzed using $y=\log_2(x)$, the RB estimator gave results about 27%-28% too high whereas the Alt and Dz estimators gave

results that were -10% to -11% too low (Table 2, Panel C). With $\sigma = 0.5$ and the difference from the baseline scenario being β_{linear} set to 0.5, the results were identical to those for the baseline scenario (Table 2, Panel D). When β_{linear} was set to 10, the results were virtually the same as the baseline scenario (Table 2, Panel E).

Application to Real Data

We identified nine published analyses of data for which the raw data were publicly available and that met our criteria for selection (Table S1; as noted above, we counted the Xu et al., 2020 study twice because they presented results for a continuous and a dichotomous outcome in relation to the same exposure) [9,11–17]. The results of our re-analyses were generally the same order of magnitude as those originally published (Table S1). The specific finding that we used in the analysis and its location in the original publication are listed in Supplementary Material Table S2, as are the quartiles of the exposure distributions, which were estimated in some cases as indicated by table footnotes. We identified six published analyses of data where the original authors presented regression results using exposure with and without a log-transformation (Table S3) [7,18–22]. Five of these were included in the assessment of validity by Dzierlenga et al. (2020) [8]. Among the fifteen example studies, a variety of outcomes and exposure variables were examined, though in two thirds of the studies the exposure was a perfluoroalkyl substance (either PFOA or PFOS).

When β was re-expressed as if it had been fit to untransformed exposure data, the proportional difference from the fitted coefficient ranged from -0.42 to 19.04 (Table 3, top half). When β was re-expressed as if it had been fit to log-transformed exposure data, the proportional difference from the observed β ranged from -0.95 to 2.44 (Table 3, bottom half). The re-expressed and fitted βs were more frequently within an order of magnitude ($-0.9 < p\Delta < 9$) using a re-expression of β to log(x) basis rather than the re-expression of β to untransformed x.

In the comparison of results across re-expression methods, when re-expressing results to the untransformed exposure basis, the proportional difference from the fitted β was, for eleven of the studies, similar across methods (top half of Table 3). These were studies where the median exposure was > 4 units. For the Lee et al. (2020) and the two Xu et al. (2020) results [9,12], however, RB estimator performed better than the other two methods [6]. In these three instances, the median of the exposure variable was less than one, which was not the case for the other studies (see Supplementary Materials Table S2). The median proportional difference from the fitted value was: RB, 0.22; Alt, 0.78, and Dz, 0.77. When re-expressing results to log(exposure) basis (bottom half of Table 3), the RB estimator generally gave results that differed from the other two estimators and the median proportional difference from the fitted value had a median closer to zero for the RB than for the other two estimators (-0.21 vs. -0.44), but the range in proportional differences was greater (RB, -0.81, 2.44; Alt, -0.94, 0.66; Dz, -0.95, 0.69).

Our results for Odebeatu et al. (2019) and Pilkerton et al. (2018) were the ones with the greatest discrepancy between the re-expressed β coefficients and the β fitted to the untransformed exposure [13,15](see top half of Table 3). This discrepancy suggested that there may have been observations that were influential, and that the influence was affected by whether the exposure had been log-transformed. Thus, we conducted an analysis of whether exclusion of influential points affected the accuracy of the re-expression. For comparison, similar analyses were conducted using the data from Cheang et al. (2021) and Xu et al., dichotomous outcome (2020), which showed smaller relative differences between the re-expressed and fitted βs. The analyses with and without the inclusion of especially influential points in the real data sets showed that the accuracy of the re-expression estimators was affected by their exclusion (Table 4). The proportional difference of re-expressed β from β fitted to untransformed x was affected by the influential points more so for Odebeatu et al. (2019) and Pilkerton et al., (2019) than for Cheang et al. (2021) and Xu et al. (2020). However, the direction of bias

due to exclusion was not consistent. The proportional difference of re-expressed β from β fitted to log(x) tended to depend less on exclusion.

Discussion

In the simulations, the bias in each of the three estimators was evaluated in relation to the median of the exposure variable, the skewness in the exposure variable, the transformation of exposure in the model generating the data (whether it was based on untransformed x, $\log_e(x)$, or $\log_2(x)$), and the β in the model generating the data. The simulations suggested there were settings in which each estimator performed well (e.g., ± 10% of the fitted regression slope), settings where the estimator performed somewhat well (e.g., ± 30% of the fitted regression slope), and settings where the estimator performed poorly (e.g., 200% greater than the fitted regression slope). Based on the simulations alone, it appeared that in the right setting, one may be able to choose an estimator that would be expected to work best, possibly with an acceptable degree of error.

Some of the results from the simulation seemed to agree with what the real data examples showed. For example, the Alt and Dz estimators performed about equally well in both the simulation and the real data examples. In the simulations the sign of the error term (pΔ) was usually positive when re-expressing to β for untransformed exposure, and negative when re-expressing to β for log(x), in general agreement with the real data examples. In the simulation the success of the estimators varied with the skewness of the exposure distribution. Among the 15 real example datasets, the 50th percentile of sigma (parameter for skewness in the exposure distribution) was 0.58, a little larger than in our baseline simulation, but one study (Odebeatu et al. 2019) had a sigma of 1.26 (Table S2). In our analysis of the Odebeatu et al. data (re-expressing to untransformed exposure), the estimators performed poorly, agreeing with the simulation. The next largest sigma in the real datasets was 0.83 (Lee et al. 2020), and while the Alt and Dz estimators fared poorly for that study, the RB was better, which was

consistent with the simulation (the median exposure in Lee et al. was 0.24). However, in another of the real data examples where the estimators performed poorly (Pilkerton et al., 2018), the sigma of the exposure distribution was 0.55, suggesting that the poor performance was not due to marked skewness. The comparison of the simulated and real data results makes it clear that predicting the degree of error in the re-expression estimator for a specific study cannot be done accurately based on descriptors such as median of exposure and a measure of its skewness. For a group of studies it may be possible to predict the general degree of error, recognizing that one or more individual studies might be re-expressed with substantially more error than predicted.

In the context of an exposure with a skewed distribution, the simulated and real data examples indicated that regardless of which of the three methods of re-expressing β were used, the results were likely to be biased. By analyzing simulated data Rodriguez-Barranco et al. (2017) recognized the importance of skewness, though the degree of skewness in their simulations was not specified. They also reported the direction of the re-expression appeared to be informative about the magnitude of bias. We identified additional factors that can have a large effect on the success of the re-expression, e.g., influential observations, and for the RB estimator, the median exposure concentration. In our example data and simulation, as a general rule, when re-expressing β from log to untransformed basis the estimated regression coefficient was likely to be larger than the actual regression coefficient. For the re-expression method proposed by Steenland et al. (2018), apparently it was assumed that if an exposure distribution had an upper bound near 10 units, their empirical re-expression method would be sufficiently accurate [7]. Our results suggested that the range of exposure was predictive of the validity of the re-expression only for the RB estimator. Using a re-expression method similar to Steenland et al.'s [8], Dzierlenga et al. (2020) conducted a validation study, using only five studies. Three of the studies in that validation had the least bias in the present study; their relatively small number of validation studies may have led to an overly optimistic appraisal of the method.

In the real data examples, our results showed a smaller median bias with the Rodriguez-Barranco et al. estimator compared with the other two methods. When converting β to untransformed exposure basis, and the median exposure is greater than four, however, the methods performed about the same. When converting β to log exposure basis, the median bias was smallest with the RB estimator, possibly reflecting the findings from the simulation showing that the Alt and Dz methods tend to underestimate the fitted beta, whereas the RB method tends to overestimate it.

Conclusions

The results of this assessment of validity have implications for meta-analysts considering or using these re-expression methods. Sensitivity analyses to evaluate whether overall findings substantially change after assuming bias in the estimated coefficients may be worthwhile. If the results requiring re-expression were neither extreme nor precise, the results of the meta-analysis may not be sensitive to the use of re-expression. Similarly, if the meta-analysis exhibits substantial heterogeneity, it is possible that the bias due to re-expression is relatively small compared with the overall variation in results.

Expanding the number and variety of studies included in this analysis of validity has provided insights into interpretation of re-expressed regression coefficients. Further expansion of the exercise by including additional studies may allow identification of specific predictors of bias and might best be done with access to raw data. Even so, when faced with a decision about whether to re-express a given finding, such an expanded study may still not signal the settings in which a specific re-expression is likely to be especially biased. On the other hand, an expanded study may be better able to define the likely range of error.

The bias due to re-expression is affected by the degree of skewness in the exposure variable, the direction of the re-expression, influential observations, and in some case, the median exposure. The re-

expression approach may be useful in selected settings, but it will result in bias with a magnitude and sign that is hard to predict.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data needed to replicate this analysis can be found in the supplemental materials for this publication.

Competing interests

Despite funding the research as noted below, 3M did not influence the research in any way, and encouraged publication in a peer-reviewed journal. The authors have no other conflicts to declare.

Funding

Support for this project (Linakis, Van Landingham, Longnecker) came from 3M to Ramboll.

Authors' contributions

Dr. Linakis ran the re-expression algorithm on the examples and simulation datasets, conducted the analysis of factors that determine validity, computed results using the other re-expression methods, and contributed to the draft of the manuscript.

Ms. Van Landingham replicated the analysis of the example NHANES studies, conducted the influence analyses, and contributed to the draft of the manuscript.

Dr. Longnecker conceived of the study, designed it, obtained funding for it, and drafted the manuscript.

Acknowledgments

Dr. Michael W. Dzierlenga provided critical comments on an early draft of the manuscript.

References

1. Deeks JJ, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. In: Cochrane Handbook for Systematic Reviews of Interventions [Internet]. John Wiley & Sons, Ltd; 2019 [cited 2021 Sep 2]. p. 241–84. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119536604.ch10

2. Higgins JP, Li T, Deeks JJ. Choosing effect measures and computing estimates of effect. In: Cochrane Handbook for Systematic Reviews of Interventions [Internet]. John Wiley & Sons, Ltd; 2019 [cited 2021 Sep 2]. p. 143–76. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119536604.ch6

3. Negri E, Metruccio F, Guercio V, Tosti L, Benfenati E, Bonzi R, et al. Exposure to PFOA and PFOS and fetal growth: a critical merging of toxicological and epidemiological data. Crit Rev Toxicol. 2017 Jul;47(6):482–508.

4. Ye X, Kong W, Zafar MI, Chen L-L. Serum triglycerides as a risk factor for cardiovascular diseases in type 2 diabetes mellitus: a systematic review and meta-analysis of prospective studies. Cardiovasc Diabetol. 2019 Apr 15;18(1):48.

5. McCabe JJ, O'Reilly E, Coveney S, Collins R, Healy L, McManus J, et al. Interleukin-6, C-reactive protein, fibrinogen, and risk of recurrence after ischaemic stroke: Systematic review and meta-analysis. Eur Stroke J. 2021 Mar;6(1):62–71.

6. Rodríguez-Barranco M, Tobías A, Redondo D, Molina-Portillo E, Sánchez MJ. Standardizing effect size from linear regression models with log-transformed variables for meta-analysis. BMC Med Res Methodol. 2017 Mar 17;17(1):44.

7. Steenland K, Barry V, Savitz D. Serum Perfluorooctanoic Acid and Birthweight: An Updated Meta-analysis With Bias Analysis. Epidemiol Camb Mass. 2018 Nov;29(6):765–76.

8. Dzierlenga MW, Crawford L, Longnecker MP. Birth weight and perfluorooctane sulfonic acid: a random-effects meta-regression analysis. Environ Epidemiol Phila Pa. 2020 Jun;4(3):e095.

9. Xu C, Liang J, Xu S, Liu Q, Xu J, Gu A. Increased serum levels of aldehydes are associated with cardiovascular disease and cardiovascular risk factors in adults. J Hazard Mater. 2020 Dec 5;400:123134.

10. Belsley DA, Kuh E, Welsch RE. Regression diagnostics: identifying influential data and sources of collinearity. New York: Wiley; 1980. 292 p. (Wiley series in probability and mathematical statistics).

11. Bulka CM, Avula V, Fry RC. Associations of exposure to perfluoroalkyl substances individually and in mixtures with persistent infections: Recent findings from NHANES 1999-2016. Environ Pollut Barking Essex 1987. 2021 Apr 15;275:116619.

12. Lee S, Min J-Y, Min K-B. Female Infertility Associated with Blood Lead and Cadmium Levels. Int J Environ Res Public Health. 2020 Mar 10;17(5):E1794.

13. Odebeatu CC, Taylor T, Fleming LE, J Osborne N. Phthalates and asthma in children and adults: US NHANES 2007-2012. Environ Sci Pollut Res Int. 2019 Sep;26(27):28256–69.

14. Stein CR, McGovern KJ, Pajak AM, Maglione PJ, Wolff MS. Perfluoroalkyl and polyfluoroalkyl substances and indicators of immune function in children aged 12-19 y: National Health and Nutrition Examination Survey. Pediatr Res. 2016 Feb;79(2):348–57.

15. Pilkerton CS, Hobbs GR, Lilly C, Knox SS. Rubella immunity and serum perfluoroalkyl substances: Sex and analytic strategy. PloS One. 2018;13(9):e0203330.

16. Cheang I, Liao S, Zhu X, Lu X, Zhu Q, Yao W, et al. Association of acrylamide hemoglobin biomarkers with serum lipid levels in general US population: NHANES 2013-2016. Ecotoxicol Environ Saf. 2021 May;214:112111.

17. Abraham K, Mielke H, Fromme H, Völkel W, Menzel J, Peiser M, et al. Internal exposure to perfluoroalkyl substances (PFASs) and biological markers in 101 healthy 1-year-old children: associations between levels of perfluorooctanoic acid (PFOA) and vaccine response. Arch Toxicol. 2020 Jun;94(6):2131–47.

18. Apelberg BJ, Witter FR, Herbstman JB, Calafat AM, Halden RU, Needham LL, et al. Cord serum concentrations of perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA) in relation to weight and size at birth. Environ Health Perspect. 2007 Nov;115(11):1670–6.

19. Washino N, Saijo Y, Sasaki S, Kato S, Ban S, Konishi K, et al. Correlations between prenatal exposure to perfluorinated chemicals and reduced fetal growth. Environ Health Perspect. 2009 Apr;117(4):660–7.

20. Hamm MP, Cherry NM, Chan E, Martin JW, Burstyn I. Maternal exposure to perfluorinated acids and fetal growth. J Expo Sci Environ Epidemiol. 2010 Nov;20(7):589–97.

21. Chen M-H, Ha E-H, Wen T-W, Su Y-N, Lien G-W, Chen C-Y, et al. Perfluorinated compounds in umbilical cord blood and adverse birth outcomes. PloS One. 2012;7(8):e42474.

22. Darrow LA, Stein CR, Steenland K. Serum perfluorooctanoic acid and perfluorooctane sulfonate concentrations in relation to birth outcomes in the Mid-Ohio Valley, 2005-2010. Environ Health Perspect. 2013 Oct;121(10):1207–13.

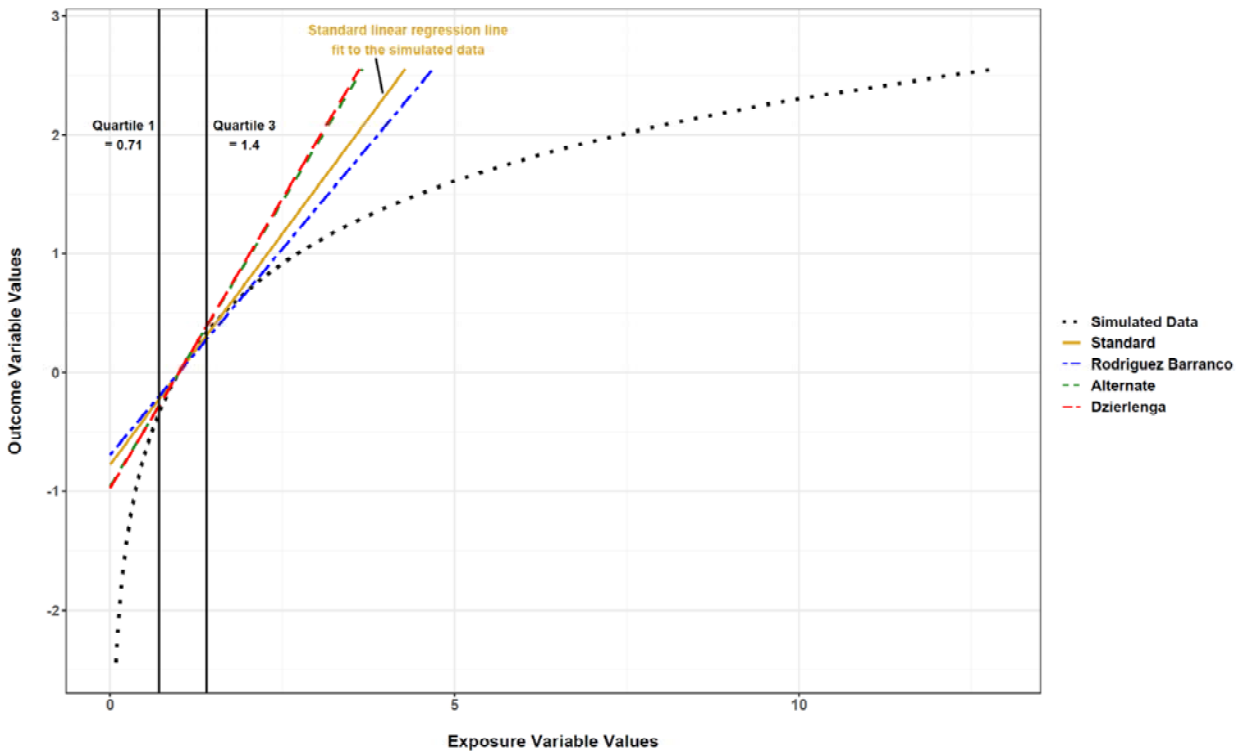

Figure 1: Simulated data (y = ln(x), dotted black curve) versus 3 methods for re-expressing beta values. Vertical lines (x-values shown on plot) indicate the first and third quartiles of the x-values. Solid (gold) line represents the line with the fitted slope (β_F) and intercept -β_F. Other dashed lines represent lines with slope $\beta_{re\text{-}expressed}$ and intercept -$\beta_{re\text{-}expressed}$ using the corresponding method noted in the plot.

Table 1. Estimated and fitted regression coefficients and the proportional difference of the estimated regression coefficients compared with the fitted coefficient (pΔ). The independent variable was x drawn at random from a $\log_e$normal distribution (10^6 draws). Results shown according to the function that generated the data, the mean of the distribution of $\log_e(x)$, the standard deviation of $\log_e(x)$, and the coefficient in the function that generated the data.

Panel A

Function that generated the data	Log_e distribution μ	Log_e distribution σ	β in function that generated the data
$y = \beta \cdot \log_e(x)$	$\mu = \log_e$(median of x shown below)	0.5	1

				Estimated regression coefficient according to method of re-expression			Fitted regression coefficient in model $y = \alpha + \beta_F \cdot x + \varepsilon$	Proportional Δ according to method of re-expression[a]		
Example number	Median of x	25th %ile x	75th %ile x	RB[b]	Alt[c]	Dz[d]	Fitted beta (β_F)[e]	RB	Alt	Dz
1	0.25	0.18	0.35	1.61	3.84	3.90	3.10	-0.48	0.24	0.26
2	0.5	0.36	0.70	1.10	1.92	1.95	1.55	-0.29	0.24	0.26
3	1	0.71	1.40	0.69	0.96	0.97	0.78	-0.11	0.23	0.25
4	2	1.43	2.80	0.41	0.48	0.49	0.39	0.04	0.24	0.26
5	4	2.85	5.61	0.22	0.24	0.24	0.19	0.15	0.24	0.26
6	8	5.71	11.2	0.12	0.12	0.12	0.10	0.22	0.24	0.26
7	16	11.4	22.4	0.06	0.06	0.06	0.05	0.25	0.24	0.26
8	32	22.8	44.8	0.03	0.03	0.03	0.02	0.27	0.23	0.25

Panel B

Function that generated the data	Log_e distribution μ	Log_e distribution σ	β in function that generated the data
$y = \beta \cdot \log_e(x)$	$\log_e$ of median shown below	1.0	1

				Estimated regression coefficient according to method of re-expression			Fitted regression coefficient in model $y = \alpha + \beta_F \cdot x + \varepsilon$	Proportional Δ according to method of re-expression[a]		
Example number	Median of x	25th %ile x	75th %ile x	RB[b]	Alt[c]	Dz[d]	Fitted beta (β_F)[e]	RB	Alt	Dz
1	0.25	0.13	0.49	1.61	3.84	3.62	1.39	0.15	1.75	1.60
2	0.5	0.25	0.98	1.1	1.92	1.81	0.71	0.56	1.72	1.56
3	1	0.51	1.97	0.69	0.96	0.90	0.35	0.97	1.72	1.56
4	2	1.02	3.92	0.41	0.48	0.45	0.18	1.30	1.72	1.57
5	4	2.03	7.87	0.22	0.24	0.23	0.09	1.55	1.74	1.58
6	8	4.07	15.7	0.12	0.12	0.11	0.04	1.68	1.73	1.58
7	16	8.14	31.4	0.06	0.06	0.06	0.02	1.74	1.72	1.56
8	32	16.3	62.7	0.03	0.03	0.03	0.01	1.77	1.70	1.55

(cont.)

Table 1, cont.

Panel C

Function that generated the data	Log_e distribution μ	Log_e distribution σ	β in function that generated the data
$y = \beta \cdot \log_2(x)$	$\log_e$ of median shown below	0.5	1

				Estimated regression coefficient according to method of re-expression			Fitted regression coefficient in model $y = \alpha + \beta_F \cdot x + \varepsilon$	Proportional Δ according to method of re-expression[a]		
Example number	Median of x	25th %ile x	75th %ile x	RB[b]	Alt[c]	Dz[d]	Fitted beta (β_F)[e]	RB	Alt	Dz
1	0.25	0.18	0.35	2.32	5.66	5.62	4.47	-0.48	0.27	0.26
2	0.5	0.36	0.70	1.58	2.83	2.81	2.24	-0.29	0.26	0.26
3	1	0.71	1.40	1.00	1.41	1.41	1.12	-0.11	0.26	0.25
4	2	1.43	2.80	0.59	0.71	0.70	0.56	0.04	0.26	0.26
5	4	2.85	5.61	0.32	0.35	0.35	0.28	0.15	0.26	0.26
6	8	5.71	11.2	0.17	0.18	0.18	0.14	0.22	0.26	0.26
7	16	11.4	22.4	0.09	0.09	0.09	0.07	0.25	0.26	0.26
8	32	22.8	44.8	0.04	0.04	0.04	0.04	0.27	0.26	0.25

Panel D

Function that generated the data	Log_e distribution μ	Log_e distribution σ	β in function that generated the data
$y = \beta \cdot \log_e(x)$	$\log_e$ of median shown below	0.5	0.5

				Estimated regression coefficient according to method of re-expression			Fitted regression coefficient in model $y = \alpha + \beta_F \cdot x + \varepsilon$	Proportional Δ according to method of re-expression[a]		
Example number	Median of x	25th %ile x	75th %ile x	RB[b]	Alt[c]	Dz[d]	Fitted beta (β_F)[e]	RB	Alt	Dz
1	0.25	0.18	0.35	0.8	1.92	1.95	1.55	-0.48	0.24	0.26
2	0.5	0.36	0.70	0.55	0.96	0.97	0.78	-0.29	0.24	0.26
3	1	0.71	1.40	0.35	0.48	0.49	0.39	-0.11	0.23	0.25
4	2	1.43	2.80	0.20	0.24	0.24	0.19	0.04	0.24	0.26
5	4	2.85	5.61	0.11	0.12	0.12	0.10	0.15	0.24	0.26
6	8	5.71	11.2	0.06	0.06	0.06	0.05	0.22	0.24	0.26
7	16	11.4	22.4	0.03	0.03	0.03	0.02	0.25	0.24	0.26
8	32	22.8	44.8	0.02	0.01	0.02	0.01	0.27	0.23	0.25

(cont.)

Table 1, cont.

Panel E

Function that generated the data	Log_e distribution μ	Log_e distribution σ	β in function that generated the data
$y = \beta \cdot \log_e(x)$	$\log_e$ of median shown below	0.5	10

				Estimated regression coefficient according to method of re-expression			Fitted regression coefficient in model $y = \alpha + \beta_F \cdot x + \varepsilon$	Proportional Δ according to method of re-expression[a]		
Example number	Median of x	25th %ile x	75th %ile x	RB[b]	Alt[c]	Dz[d]	Fitted beta (β_F)[e]	RB	Alt	Dz
1	0.25	0.18	0.35	16.1	38.4	39.01	30.99	-0.48	0.24	0.26
2	0.5	0.36	0.70	11.0	19.2	19.49	15.53	-0.29	0.24	0.25
3	1	0.71	1.40	6.93	9.69	9.74	7.78	-0.11	0.23	0.25
4	2	1.43	2.80	4.06	4.80	4.88	3.88	0.04	0.24	0.26
5	4	2.85	5.61	2.23	2.40	2.44	1.94	0.15	0.24	0.26
6	8	5.71	11.2	1.18	1.20	1.22	0.97	0.22	0.24	0.26
7	16	11.4	22.4	0.61	0.60	0.61	0.49	0.25	0.24	0.26
8	32	22.8	44.8	0.31	0.30	0.30	0.24	0.27	0.23	0.25

[a] Proportional difference of the estimated regression coefficient compared with the fitted coefficient, $\beta_E/\beta_F - 1$.
[b] RB, Rodriquez-Barranco et al. estimator, modified (see text).
[c] Alternate estimator (see text).
[d] Dzierlenga et al. estimator (see text).
[e] The R^2 for all examples in Panels A, C, D, E, F, H, I and J rounded to 0.88; for Panels B and G, to 0.58.

Table 2. Estimated and fitted regression coefficients and the proportional difference of the estimated regression coefficients compared with the fitted coefficient (pΔ). The independent variable was x drawn at random from a $\log_e$normal distribution (10^6 draws). Results shown according to the function that generated the data, the mean of the distribution of $\log_e(x)$, the standard deviation of $\log_e(x)$, and the coefficient in the function that generated the data.

Panel A

Function that generated the data	Log_e distribution μ	Log_e distribution σ	β in function that generated the data
$y = \beta \cdot x$	$\log_e$ of median shown below	0.5	1

				Estimated regression coefficient according to method of re-expression			Fitted regression coefficient in model $y = \alpha + \beta_F \cdot \log_e(x) + \varepsilon$	Proportional Δ according to method of re-expression[a]		
Example number	Median of x	25th %ile x	75th %ile x	RB[b]	Alt[c]	Dz[d]	Fitted beta (β_F)[e]	RB	Alt	Dz
1	0.25	0.18	0.35	0.43	0.26	0.25	0.28	0.51	-0.08	-0.11
2	0.5	0.36	0.7	0.86	0.52	0.51	0.57	0.52	-0.08	-0.10
3	1	0.71	1.4	1.72	1.04	1.02	1.13	0.52	-0.08	-0.10
4	2	1.43	2.8	3.43	2.08	2.03	2.27	0.52	-0.08	-0.10
5	4	2.85	5.61	6.87	4.17	4.06	4.54	0.51	-0.08	-0.10
6	8	5.71	11.2	13.7	8.34	8.12	9.08	0.51	-0.08	-0.10
7	16	11.4	22.4	27.5	16.7	16.3	18.1	0.52	-0.08	-0.10
8	32	22.8	44.8	55.0	33.4	32.5	36.2	0.52	-0.08	-0.10

Panel B

Function that generated the data	Log_e distribution μ	Log_e distribution σ	β in function that generated the data
$y = \beta \cdot x$	$\log_e$ of median shown below	1.0	1

				Estimated regression coefficient according to method of re-expression			Fitted regression coefficient in model $y = \alpha + \beta_F \cdot \log_e(x) + \varepsilon$	Proportional Δ according to method of re-expression[a]		
Example number	Median of x	25th %ile x	75th %ile x	RB[b]	Alt[c]	Dz[d]	Fitted beta (β_F)[e]	RB	Alt	Dz
1	0.25	0.13	0.49	0.43	0.26	0.27	0.41	0.04	-0.37	-0.36
2	0.5	0.25	0.98	0.86	0.52	0.53	0.83	0.04	-0.37	-0.36
3	1	0.51	1.97	1.72	1.04	1.06	1.64	0.05	-0.37	-0.35
4	2	1.02	3.92	3.43	2.08	2.12	3.3	0.04	-0.37	-0.36
5	4	2.03	7.87	6.87	4.17	4.25	6.61	0.04	-0.37	-0.36
6	8	4.07	15.7	13.7	8.34	8.50	13.2	0.04	-0.37	-0.36
7	16	8.14	31.4	27.5	16.7	17.0	26.4	0.04	-0.37	-0.36
8	32	16.3	62.7	55.0	33.3	34.0	52.5	0.05	-0.37	-0.35

(cont.)

Table 2, cont.

Panel C

Function that generated the data	Log_e distribution μ	Log_e distribution σ	β in function that generated the data
$y = \beta \cdot x$	$\log_e$ of median shown below	0.5	1

				Estimated regression coefficient according to method of re-expression			Fitted regression coefficient in model $y = \alpha + \beta_F \cdot \log_2(x) + \varepsilon$	Proportional Δ according to method of re-expression[a]		
Example number	Median of x	25th %ile x	75th %ile x	RB[b]	Alt[c]	Dz[d]	Fitted beta (β_F)[e]	RB	Alt	Dz
1	0.25	0.18	0.35	0.25	0.18	0.18	0.2	0.27	-0.10	-0.11
2	0.5	0.36	0.70	0.50	0.35	0.35	0.39	0.27	-0.10	-0.10
3	1	0.71	1.40	1.00	0.71	0.7	0.78	0.28	-0.10	-0.10
4	2	1.43	2.80	2.00	1.41	1.41	1.57	0.27	-0.10	-0.10
5	4	2.85	5.61	4.00	2.83	2.81	3.15	0.27	-0.10	-0.10
6	8	5.71	11.2	8.00	5.66	5.63	6.29	0.27	-0.10	-0.10
7	16	11.4	22.4	16.0	11.3	11.3	12.6	0.27	-0.10	-0.10
8	32	22.8	44.8	32.0	22.6	22.5	25.1	0.28	-0.10	-0.10

Panel D

Function that generated the data	Log_e distribution μ	Log_e distribution σ	β in function that generated the data
$y = \beta \cdot x$	$\log_e$ of median shown below	0.5	0.5

				Estimated regression coefficient according to method of re-expression			Fitted regression coefficient in model $y = \alpha + \beta_F \cdot \log_e(x) + \varepsilon$	Proportional Δ according to method of re-expression[a]		
Example number	Median of x	25th %ile x	75th %ile x	RB[b]	Alt[c]	Dz[d]	Fitted beta (β_F)[e]	RB	Alt	Dz
1	0.25	0.18	0.35	0.21	0.13	0.13	0.14	0.51	-0.08	-0.11
2	0.5	0.36	0.7	0.43	0.26	0.25	0.28	0.52	-0.08	-0.10
3	1	0.71	1.4	0.86	0.52	0.51	0.57	0.52	-0.08	-0.10
4	2	1.43	2.8	1.72	1.04	1.01	1.13	0.52	-0.08	-0.10
5	4	2.85	5.61	3.44	2.08	2.03	2.27	0.51	-0.08	-0.10
6	8	5.71	11.2	6.87	4.17	4.06	4.54	0.51	-0.08	-0.10
7	16	11.4	22.4	13.8	8.34	8.13	9.07	0.52	-0.08	-0.10
8	32	22.8	44.8	27.5	16.7	16.3	18.1	0.52	-0.08	-0.10

(cont.)

Table 2, cont.

Panel E

Function that generated the data	Log_e distribution μ	Log_e distribution σ	β in function that generated the data
$y = \beta \cdot x$	log_e of median shown below	0.5	10

				Estimated regression coefficient according to method of re-expression			Fitted regression coefficient in model $y = \alpha + \beta_F \cdot \log_e(x) + \varepsilon$	Proportional Δ according to method of re-expression[a]		
Example number	Median of x	25th %ile x	75th %ile x	RB[b]	Alt[c]	Dz[d]	Fitted beta (β_F)[e]	RB	Alt	Dz
1	0.25	0.18	0.35	4.29	2.6	2.54	2.84	0.51	-0.08	-0.11
2	0.5	0.36	0.7	8.59	5.21	5.08	5.67	0.52	-0.08	-0.10
3	1	0.71	1.4	17.2	10.4	10.2	11.3	0.52	-0.08	-0.10
4	2	1.43	2.8	34.3	20.8	20.3	22.7	0.52	-0.08	-0.10
5	4	2.85	5.61	68.7	41.7	40.6	45.4	0.51	-0.08	-0.10
6	8	5.71	11.2	137	83.4	81.3	90.8	0.51	-0.08	-0.10
7	16	11.4	22.4	275	167	163	181	0.52	-0.08	-0.10
8	32	22.8	44.8	550	333	325	362	0.52	-0.08	-0.10

[a] Proportional difference of the estimated regression coefficient compared with the fitted coefficient, $\beta_E/\beta_F - 1$.
[b] RB, Rodriquez-Barranco et al. estimator, modified (see text).
[c] Alternate estimator (see text).
[d] Dzierlenga et al. estimator (see text).
[e] The R^2 for all examples in Panels A, C, D, E, F, H, I and J rounded to 0.88; for Panels B and G, to 0.58.

Table 3: Comparison of fitted and re-expressed β coefficients and proportional difference in β for three methods of re-expression

First author, year	Analysis of raw data (fitted β)	Modified Rodriguez-Barranco et al.[a]	pΔ[b]	Alternative re-expression method[c]	pΔ[b]	Method presented in this report	pΔ[b]
Abraham, 2020	-0.0636 $\log_e$/ ng·ml^{-1}	-0.03788	-0.40	-0.03841	-0.40	-0.03684	-0.42
Apelberg, 2007	-12.9 g/(ng/ml)	-12.58	-0.02	-13.24	0.03	-13.25	0.03
Bulka, 2021	0.015 $\log_e$(OR)/ ng·ml^{-1}	0.0462	2.08	0.0531	2.54	0.0512	2.41
Cheang, 2021	0.192 mg·dl/pmol·g^{-1}	0.3552	0.85	0.3526	0.84	0.3522	0.83
Chen, 2012	-11.3 g/(ng/ml)	-17.15	0.52	-17.80	0.58	-17.86	0.58
Darrow, 2013	-2.3 g/(ng/ml)	-2.01	-0.12	-2.00	-0.13	-2.03	-0.12
Hamm, 2010	1.5 g/(ng/ml)	3.78	1.52	3.85	1.57	3.92	1.62
Lee, 2020	1.444 $\log_e$(OR)/µg·dl^{-1}	1.3931	-0.04	3.465	1.40	3.2953	1.28
Odebeatu, 2019	6.86x10^{-4} $\log_e$(OR)/ng·ml^{-1}	1.37x10^{-2}	19.04	1.16x10^{-2}	15.86	1.22x10^{-2}	16.82
Pilkerton, 2018	-0.0049 $\log_e$/ng·ml^{-1}	-0.02805	4.73	-0.03059	5.24	-0.03023	5.17
Steenland, 2009	0.00105 $\log_e$/ ng·ml^{-1}	0.001285	0.22	0.001264	0.20	0.001273	0.21
Stein, 2016	-0.0039 $\log_e$/ng·ml^{-1}	-0.00693	0.78	-0.00694	0.78	-0.00690	0.77
Washino, 2009	-10.94 g/(ng/ml)	-11.367	0.04	-10.054	-0.08	-12.073	0.10
Xu, 2020a	0.513 $\log_e$(OR)/ ng·ml^{-1}	0.5734	0.12	1.0071	0.96	0.9616	0.87
Xu, 2020b	29.3 mg·dl/ng·ml^{-1}	28.70	-0.02	50.41	0.72	48.13	0.64
Median			0.22		0.78		0.77
25th percentile			-0.02		0.12		0.16
75th percentile			1.19		1.49		1.45
Minimum			-0.40		-0.40		-0.42
Maximum			19.04		15.86		16.82
Abraham, 2020	-0.3884 $\log_e$/$\log_2$	-0.90948	1.34	-0.64310	0.66	-0.65479	0.69
Apelberg, 2007	-69 g/$\log_e$(ng/ml)	-110.829	0.61	-67.2	-0.03	-66.1	-0.04
Bulka, 2021	0.104 $\log_e$(OR)/$\log_2$	0.04155	-0.60	0.0294	-0.72	0.0298	-0.71
Cheang, 2021	9.65 mg·dl/$\log_2$	7.4304	-0.23	5.254	-0.46	5.223	-0.46
Chen, 2012	-110.2 g/$\log_e$(ng/ml)	-115.335	0.05	-69.95	-0.37	-68.65	-0.38
Darrow, 2013	-29 g/$\log_e$(ng/ml)	-54.9335	0.89	-33.3	0.15	-32.5	0.12
Hamm, 2010	31.3 g/$\log_e$(ng/ml)	20.1039	-0.36	12.19	-0.61	11.86	-0.62
Lee, 2020	0.588 $\log_e$(OR)/$\log_2$	0.34656	-0.41	0.2451	-0.58	0.2506	-0.57
Odebeatu, 2019	0.405 $\log_e$(OR)/$\log_{10}$	0.07594	-0.81	0.0240	-0.94	0.0214	-0.95
Pilkerton, 2018	-0.093 $\log_e$/$\log_2$	-0.02107	-0.77	-0.0149	-0.84	-0.0149	-0.84
Steenland, 2009	0.0266 $\log_e$/$\log_e$	0.036445	0.37	0.02210	-0.17	0.02163	-0.19
Stein, 2016	-0.109 $\log_e$/$\log_2$	-0.08658	-0.21	-0.0612	-0.44	-0.0610	-0.44
Washino, 2009	-148.8 g/$\log_{10}$(ng/ml)	-511.992	2.44	-161.91	0.09	-133.29	-0.10
Xu, 2020a	0.371 $\log_e$(OR)/$\log_2$	0.267273	-0.28	0.1890	-0.49	0.1929	-0.48
Xu, 2020b	18.57 mg·dl/ $\log_2$	15.2653	-0.18	10.794	-0.42	11.016	-0.41
Median			-0.21		-0.44		-0.44
25th percentile			-0.39		-0.60		-0.60
75th percentile			0.49		-0.10		-0.15
Minimum			-0.81		-0.94		-0.95
Maximum			2.44		0.66		0.69

(cont.)

[a] The re-expression formula originally presented by Rodriguez-Barranco et al. (2017) was based on an observed β and a mean value of x, the exposure variable. In the table above, however, we used the median of x, because when the distribution of x is skewed, a median is more likely to be presented than a mean. Using the notation of Rodriguez-Barranco et al., k = base of log transformation used; c = 1.
[b] Proportional difference between β in column to the left compared with the one from the analysis of raw data, calculated using the same method as in previous table ((beta in column to left – beta from analysis of raw data)/beta from analysis of raw data). Note that the β in column to the left was calculated using β with different units in denominator than for the raw data analysis shown in the table.
[c] Let the log unit increment I in untransformed units = $b^{(\log_b(\text{median})+0.5)} - b^{(\log_b(\text{median})-0.5)}$, where b = 2, e, or 10, depending on the base. For observed β_o with units of $\Delta y/\Delta \log_b(x)$, to get re-expressed β_r with units $\Delta y/\Delta x$, calculate $\beta_r = \beta_o/I$. If the units of β_o are $\Delta y/\Delta x$, to get β_r with units $\Delta y/\Delta \log_b(x)$, calculate $\beta_r = \beta_o \cdot I$.

Table 4.

1st author, year	Proportional difference with influential observations	Proportional difference with no influential observations
Proportional difference of re-expressed β from β fitted to untransformed x		
Odebeatu 2019	16.82	19.58
Pilkerton 2018	5.17	2.15
Cheang 2021	0.84	1.11
Xu 2020a	0.88	1.38
Proportional difference of re-expressed β from β fitted to log(x)		
Odebeatu 2019	-0.95	-0.99
Pilkerton 2018	-0.84	-0.37
Cheang 2021	-0.46	-0.11
Xu 2020a	-0.48	-0.38