

# konfound: An R Sensitivity Analysis Package to Quantify the Robustness of Causal Inferences


Sarah Narvaiz<sup>1\*</sup>, Qinyun Lin<sup>2\*</sup>, Joshua M. Rosenberg<sup>1\*¶</sup>, Kenneth A. Frank<sup>3</sup>, Spiro J. Maroulis<sup>4</sup>, Wei Wang<sup>1</sup>, and Ran Xu<sup>5</sup>

1 University of Tennessee, Knoxville, Knoxville, TN, USA 2 University of Gothenburg, Gothenburg, SE 3 Michigan State University, East Lansing, MI, USA 4 Arizona State University, Tempe, AZ, USA 5 University of Connecticut, Hartford, CT, USA ¶ Corresponding author \* These authors contributed equally.

DOI: [10.21105/joss.05779](https://doi.org/10.21105/joss.05779)

## Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: Fabian Scheipl ↗ 

## Reviewers:

- [@fartist](#)
- [@wjakethompson](#)

Submitted: 31 July 2023

Published: 26 February 2024

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Quantifying the Robustness of Inferences

Sensitivity analysis, a statistical method crucial for validating inferences across disciplines, quantifies the conditions that could alter conclusions ([Razavi et al., 2021](#)). One line of work is rooted in linear models and foregrounds the sensitivity of inferences to the strength of omitted variables ([Cinelli & Hazlett, 2019](#); [Frank, 2000](#)). A more recent approach is rooted in the potential outcomes framework for causal inference and foregrounds how hypothetical changes in a sample would alter an inference if such cases were otherwise observed ([Frank et al., 2008, 2013](#); [Frank & Min, 2007](#); [Xu et al., 2019](#)).

One sensitivity measure is the *Impact Threshold of a Confounding Variable*, or ITCV, which generates statements about the correlation of an omitted, confounding variable with both a predictor of interest and the outcome ([Frank, 2000](#)). The ITCV index can be calculated for any linear model. The *Robustness of an Inference to Replacement*, RIR, assesses how replacing a certain percentage of cases with counterfactuals of zero treatment effect could nullify an inference ([Frank et al., 2013](#)). The RIR index is more general than the ITCV index.

The sensitivity analysis techniques we describe in this paper and implement in the konfound R package differ from others in several ways. Unlike Linden et al. ([2020](#)), whose approach focuses on dichotomous outcomes and omitted variable sensitivity, our approach extends to continuous outcomes and evaluates both changes in estimates and standard errors. Oster ([2019](#)) focuses only on selection into the treatment based on unobservable variables versus observable variables necessary to nullify an estimate. The ITCV index focuses on the relationship of the unobservable to the predictor of interest and to the outcome. More generally, many others used simulation-based approaches, while our approach uses closed-form expressions to generate a single term representing sensitivity. These techniques, along with others, are reviewed and discussed (along with the ITCV and RIR approaches) in Frank et al. ([2023](#)).

We have implemented the calculation of both the ITCV and RIR indices in the konfound R package. This package is intended to provide an easy-to-use and principled set of sensitivity techniques that are suitable for a range of model and dependent variable types and use cases. Its audience is broad: primarily social scientists, but also interested individuals in other disciplines (e.g., the health sciences). This paper provides an overview of two core functions within the konfound package, each of which can calculate the ITCV and RIR indices: `konfound()` and `pkonfound()`. These functions allow users to calculate the robustness of inferences using a model estimated (with R) or using information about a model from a published study, respectively.

The konfound package is available from the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=konfound>; it can be installed via

the `install.packages("konfound")` function within R.

## Functionality

### **konfound**

This function calculates the ITCV and RIR for models fitted in R. This function currently works for linear models fitted with `lm()`, `glm()`, and `lmer()`. The output printed in the R console is the bias that must be present and the number of cases that would have to be replaced with cases for which there is no effect to nullify the inference.

#### *Example for linear models fit with `lm()`*

For this example, we will use the `concord1` dataset built into the `konfound` package. This dataset comes from a study that examines the causal mechanism behind household water conservation in a U.S. city.

We will estimate the effect of the following variables on household water consumption in 1981:

- household water consumption in 1980 (`water80`)
- household income (`income`)
- education level of household survey respondent (`educat`)
- retirement status of respondent (`retire`)
- number of individuals in household in 1980 (`peop80`)

Here is the code we use to fit a linear model using these variables:

```
m <- lm(water81 ~ water80 + income + educat + retire + peop80, data = concord1)
```

The results of the model fitting (which can be obtained by running `summary(m)` within R) indicate that all of the predictors apart from `retire` have a statistically significant effect on water consumption. In the example, we focus on the coefficient for `peop80` ( $\beta = 225.198$ ,  $SE = 28.704$ ,  $t = 7.845$ ,  $p < .001$ ).

#### *ITCV example for linear models fit with `lm()`*

Now, let's examine the robustness of the `peop80` effect by calculating the ITCV:

```
library(konfound)
```

```
konfound(m, peop80, index = "IT")
```

```
## Impact Threshold for a Confounding Variable:
## The minimum impact of an omitted variable to invalidate an inference
## for a null hypothesis of 0 effect is based on a correlation of 0.52 with
## the outcome and at 0.52 with the predictor of interest (conditioning on
## observed covariates) based on a threshold of 0.089 for statistical
## significance (alpha = 0.05).
##
## Correspondingly the impact of an omitted variable (as defined in Frank
## 2000) must be  $0.52 \times 0.52 = 0.27$  to invalidate an inference for a null
## hypothesis of 0 effect. See Frank (2000) for a description of the method.
##
## Citation:
## Frank, K. (2000). Impact of a confounding variable on the
## inference of a regression coefficient. Sociological Methods and Research,
## 29(2), 147-194
##
## For more detailed output, consider setting `to_return` to table
```

```
## To consider other predictors of interest, consider setting `test_all` to
## TRUE.
```

The output indicates that to invalidate the inference that `peop80` has an effect on `water81` using statistical significance as a threshold (e.g.,  $p = .05$ ), an omitted variable would have to be correlated at 0.520 with `peop80` and 0.520 with `water81`, conditioning on observed covariates.

#### *RIR example for linear models fit with `lm()`*

We can also examine the robustness by calculating the RIR:

```
konfound(m, peop80, index = "RIR")
```

```
## Robustness of Inference to Replacement (RIR):
## To invalidate an inference, 74.955 % of the
## estimate would have to be due to bias.
## This is based on a threshold of 56.4 for statistical
## significance (alpha = 0.05).
##
## To invalidate an inference, 372 observations would
## have to be replaced with cases for which the effect is 0 (RIR = 372).
##
## See Frank et al. (2013) for a description of the method.
##
## Citation: Frank, K.A., Maroulis, S., Duong, M., and Kelcey, B. (2013).
## What would it take to change an inference?
## Using Rubin's causal model to interpret the robustness of causal inferences.
## Education, Evaluation and Policy Analysis, 35 437-460.

## For more detailed output, consider setting `to_return` to table

## To consider other predictors of interest,
## consider setting `test_all` to TRUE.
```

The output presents two interpretations of the RIR. First, 74.956% of the estimated effect of `peop80` on `water81` would have to be attributed to bias to invalidate the inference. Equivalently, we would expect to have to replace 372 out of the 486 households (about 76%) with cases for which `peop80` had no effect to invalidate the inference. We have created guidelines for evaluating the RIR relative to the bias accounted for by observed covariates or published norms (Frank et al., 2013, 2021).

#### **pkonfound**

This function quantifies the sensitivity for analyses that have already been conducted, such as in an already-published study or from analysis carried out using other software. This function calculates how much bias there must be in an estimate to invalidate/sustain an inference, which can be interpreted as the percentage of cases that would have to be replaced (e.g., with cases for which the predictor had no effect on the outcome) to invalidate the inference. It also calculates the impact of an omitted variable necessary to invalidate/sustain an inference for a regression coefficient, where impact is defined as the correlation between the omitted variable and the focal predictor multiplied with the correlation between the omitted variable and the outcome.

#### *ITCV example for a regression analysis*

For this example, the following estimated quantities from an estimated regression model would be entered into the `pkonfound` function: unstandardized beta coefficient for the predictor of interest (`est_eff = 2`), estimated standard error (`std_err = .4`), sample size (`n_obs = 100`), and the number of covariates (`n_covariates = 3`), as follows:

```
pkonfound(2, .4, 100, 3, index = "IT")

## Impact Threshold for a Confounding Variable:
## The minimum impact of an omitted variable to invalidate an inference for
## a null hypothesis of 0 effect is based on a correlation of 0.568 with
## the outcome and at 0.568 with the predictor of interest (conditioning
## on observed covariates) based on a threshold of 0.201 for statistical
## significance (alpha = 0.05).
##
## Correspondingly the impact of an omitted variable (as defined in Frank
## 2000) must be 0.568 X 0.568 = 0.323 to invalidate an inference for a null
## hypothesis of 0 effect. See Frank (2000) for a description of the method.
##
## Citation:
## Frank, K. (2000). Impact of a confounding variable on the inference of a
## regression coefficient. Sociological Methods and Research, 29 (2), 147-194

## For other forms of output, run ?pkonfound and inspect the to_return argument

## For models fit in R, consider use of konfound().
```

#### *RIR example for a regression analysis*

We can also use the same inputs to calculate output for the RIR index:

```
pkonfound(2, .4, 100, 3, index = "RIR")

## Robustness of Inference to Replacement (RIR):
## To invalidate an inference, 60.29 % of the estimate would have to be
## due to bias. This is based on a threshold of 0.794 for statistical
## significance (alpha = 0.05).
##
## To invalidate an inference, 60 observations would have to be replaced
## with cases for which the effect is 0 (RIR = 60).
##
## See Frank et al. (2013) for a description of the method.
##
## Citation: Frank, K.A., Maroulis, S., Duong, M., and Kelcey, B. (2013).
## What would it take to change an inference?
## Using Rubin's causal model to interpret the robustness of causal inferences.
## Education, Evaluation and Policy Analysis, 35 437-460.

## For other forms of output, run ?pkonfound and inspect the to_return argument

## For models fit in R, consider use of konfound().
```

### Doing and Learning More

We have created a website including a Shiny interactive web application at <http://konfound-it.com> that can be applied to linear models, 2x2 tables resulting corresponding to treatment and control by success and failure conditions, and logistic regression models. We are also developing

extensions of the sensitivity analysis techniques described in this paper, including preserving the standard error (Frank et al., 2023) and calculating the coefficient of proportionality (Frank et al., 2022) for ITCV analyses. Functionality for designs including mediation, hazard functions, differences in difference, and regression discontinuity are also presently under development. Additional documentation on the R package and its future extensions are available at the <http://konfound-it.com> website.

## Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D220022 to Michigan State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

## References

- Cinelli, C., & Hazlett, C. (2019). Making Sense of Sensitivity: Extending Omitted Variable Bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1), 39–67. <https://doi.org/10.1111/rssb.12348>
- Frank, K. A. (2000). Impact of a Confounding Variable on a Regression Coefficient. *Sociological Methods & Research*, 29(2), 147–194. <https://doi.org/10.1177/0049124100029002001>
- Frank, K. A., Lin, Q., Maroulis, S., Dai, S., Jess, N., Lin, H., Maestrales, S., Searle, E., Tait, J., & others. (2022). Improving oster's  $\delta^*$ : Exact calculation for the coefficient of proportionality without subjective specification of a baseline model. *Ellen and Tait, Jordan, Improving Oster's  $\delta^*$ : Exact Calculation for the Coefficient of Proportionality Without Subjective Specification of a Baseline Model* (December 16, 2022). <https://doi.org/10.2139/ssrn.4305243>
- Frank, K. A., Lin, Q., Maroulis, S., Mueller, A. S., Xu, R., Rosenberg, J. M., Hayter, C. S., Mahmoud, R. A., Kolak, M., Dietz, T., & Zhang, L. (2021). Hypothetical case replacement can be used to quantify the robustness of trial results. *Journal of Clinical Epidemiology*, 134, 150–159. <https://doi.org/10.1016/j.jclinepi.2021.01.025>
- Frank, K. A., Lin, Q., Xu, R., Maroulis, S., & Mueller, A. (2023). Quantifying the robustness of causal inferences: Sensitivity analysis for pragmatic social science. *Social Science Research*, 110. <https://doi.org/10.1016/j.ssresearch.2022.102815>
- Frank, K. A., Maroulis, S. J., Duong, M. Q., & Kelcey, B. M. (2013). What would it take to change an inference? Using rubin's causal model to interpret the robustness of causal inferences. *Educational Evaluation and Policy Analysis*, 35(4), 437–460. <https://doi.org/10.3102/0162373713493129>
- Frank, K. A., & Min, K.-S. (2007). Indices of robustness for sample representation. *Sociological Methodology*, 37(1), 349–392. <https://doi.org/10.1111/j.1467-9531.2007.00186.x>
- Frank, K. A., Sykes, G., Anagnostopoulos, D., Cannata, M., Chard, L., Krause, A., & McCrory, R. (2008). Does NBPTS certification affect the number of colleagues a teacher helps with instructional matters? *Educational Evaluation and Policy Analysis*, 30(1), 3–30. <https://doi.org/10.3102/0162373707313781>
- Linden, A., Mathur, M. B., & VanderWeele, T. J. (2020). Conducting sensitivity analysis for unmeasured confounding in observational studies using e-values: The evalua package. *The Stata Journal*, 20(1), 162–175. <https://doi.org/10.1177/1536867X20909696>
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2), 187–204.

- Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J. H. A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., ... Maier, H. R. (2021). The future of sensitivity analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software*, 137. <https://doi.org/10.1016/j.envsoft.2020.104954>
- Xu, R., Frank, K. A., Maroulis, S. J., & Rosenberg, J. M. (2019). Konfound: Command to quantify robustness of causal inferences. *The Stata Journal*, 19(3), 523–550. <https://doi.org/10.1177/1536867X19874223>