

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

31 July 2023

## Quantifying the Robustness of Inferences

Statistical methods which quantify the conditions necessary to alter inferences are important to a variety of disciplines (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 (K. A. Frank 2000; Cinelli and Hazlett 2019). 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 variables or cases were otherwise observed [K. A. Frank and Min (2007); K. A. Frank et al. (2008); K. A. Frank et al. (2013); Xu et al. (2019)]. **It differs from other sensitivity analysis approaches in the variety of dependent variable types it accomodates, its internal logic, and its analytic (rather than simulation-based) approach. (K. A. Frank et al. 2023).**

We have implemented two measures from these lines of work within R via the `konfound` R package. **The audience for this R package is primarily social scientists, as well as interested individuals in other fields.** One measure is the Impact Threshold of a Confounding Variable (ITCV), which generates statements such as “to invalidate an inference of an effect, an omitted variable would have to be correlated at \_\_\_ with the predictor of interest and with the outcome” (K. A. Frank 2000). This sensitivity analysis can be calculated for any linear model. The second measure is the Robustness of an Inference to Replacement (RIR) which generates statements such as “to invalidate the inference, \_\_\_ % of the cases would have to be replaced with counterfactual cases with zero effect of the treatment” (K. A. Frank et al. 2013). The RIR represents a more generally applicable approach not limited to linear models, and is recommended for all cases that use binary outcomes (K. A. Frank et al. 2021).

## Statement of Need: The Need for an R Package

We have implemented these recent developments of sensitivity analysis for causal inferences within R via the `konfound` R package. In particular, the `konfound` package is used to calculate two robustness indices: ITCV and RIR.

This paper provides an overview of two core functions within the `konfound` package: `konfound()`, `pkonfound()`. These functions allow users to calculate the robustness of inferences using a model estimated in R in or using information about a model from a published study, respectively. The package is available from the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=konfound>; it can be installed via the standard `install.packages("konfound")` function. These two core functions within the `konfound` package are briefly summarized below along with their functionality.

## Summary and Functionality

1. **konfound**: This function is used for fitted models and calculates the ITCV and RIR. This function currently works for linear models created with `lm()`, `glm()`, and `lmer()` fitted in R. 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 invalidate 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 Hamilton's (1983) study which examines the causal mechanism behind household water conservation in a U.S. city.

We will model the impact the following variables have 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`), and number of individuals in household in 1980 (`peop80`).

```
library(konfound)

## Sensitivity analysis as described in Frank, Maroulis, Duong, and Kelcey
## (2013) and in Frank (2000).
## For more information visit http://konfound-it.com.

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

##
## Call:
## lm(formula = water81 ~ water80 + income + educat + retire + peop80,
##     data = concord1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4035.5  -453.4   -62.7   384.2  4995.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 299.7437    210.0136   1.427  0.15414
## water80      0.4943     0.0268  18.445 < 2e-16 ***
## income      22.6031     3.5023   6.454 2.62e-10 ***
## educat     -44.2578    13.4381  -3.293  0.00106 **
## retire     155.4727    96.3389   1.614  0.10721
## peop80     225.1984    28.7048   7.845 2.73e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 864.1 on 490 degrees of freedom
## Multiple R-squared:  0.6653, Adjusted R-squared:  0.6619
## F-statistic: 194.8 on 5 and 490 DF, p-value: < 2.2e-16
```

Results indicate that all variables except `retire` have a significant effect on 1981 household water consumption.

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

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

```
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 X 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 in order to invalidate the inference that `peop80` has an effect on `water81` using statistical significance as a threshold (e.g.,  $p=0.05$ ), an omitted variable would have to be correlated at 0.52 with `peop80` and 0.52 with `water81`, conditioning on observed covariates. Correspondingly, the impact of an omitted variable as defined in (K. A. Frank 2000) must be  $0.52 \times 0.52 = 0.27$ .

### *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 on evaluating the RIR relative to the bias accounted for by observed covariates or published norms (K. A. Frank et al. 2013, 2021).

2. **pkonfound**: This function quantifies the sensitivity for analyses which have already been conducted, such as in an already-published study or from analysis carried out using other software. This function calculates 1) 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 (with cases for which the predictor had no effect on the outcome) to invalidate the inference and 2) the impact of an omitted variable necessary to invalidate/sustain an inference for a regression coefficient (note: impact is defined as the correlation between omitted variable and focal predictor x correlation between omitted variable and outcome).

### *ITCV example for a regression analysis*

For this example, the following parameters from a regression analysis 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).

```
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().
```

The output indicates that in order to invalidate the inference that the predictor of interest has a greater than zero effect, an omitted variable would have to be correlated at 0.568 with the outcome and at 0.568 with the predictor of interest, conditioning on observed covariates. Correspondingly, the impact of an omitted variable (ITCV) must be  $0.568 \times 0.568 = 0.323$  to invalidate an inference for a null hypothesis of 0 effect.

## *RIR example for a regression analysis*

```
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().
```

The output indicates that to invalidate the inference, approximately 60% of the estimated effect of the predictor of interest would have to be due to bias in this model to invalidate this inference. Equivalently, 60 out of 100 cases would have to be replaced with zero effect cases in order to invalidate this inference. See (K. A. Frank et al. 2013, 2021) for guidelines on evaluating the RIR relative to the bias accounted for by observed covariates or published norms.

## Doing and Learning More

We have created a Shiny interactive web application at <http://konfound-it.com/> that can be applied to linear models, 2 x 2 tables resulting corresponding to treatment and control by success and failure conditions, and logistic regression models (functionality for designs including mediation, differences in difference, and regression discontinuity are under development). We are also developing extensions of the sensitivity analysis techniques described in this paper, including preserving the standard error (K. A. Frank et al. 2023) and calculating the coefficient of proportionality (K. Frank et al. 2022) for ITCV analyses. Additional documentation on the R package and future extensions will be available at <https://konfound-project.github.io/konfound/>.

## 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., and C. Hazlett. 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): 147–94. <https://doi.org/10.1177/0049124100029002001>.
- Frank, K. A., Q. Lin, S. Maroulis, A. S. Mueller, R. Xu, J. M. Rosenberg, C. S. Hayter, et al. 2021. “Hypothetical Case Replacement Can Be Used to Quantify the Robustness of Trial Results.” *Journal of Clinical Epidemiology* 134: 150–59. <https://doi.org/10.1016/j.jclinepi.2021.01.025>.
- Frank, K. A., Q. Lin, R. Xu, S. Maroulis, and A. Mueller. 2023. “Quantifying the Robustness of Causal Inferences: Sensitivity Analysis for Pragmatic Social Science.” *Social Science Research* 110 (February). <https://doi.org/10.1016/j.ssresearch.2022.102815>.
- Frank, K. A., S. J. Maroulis, M. Q. Duong, and B. M. Kelcey. 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): 437–60. <https://doi.org/10.3102/0162373713493129>.
- Frank, K. A., and K. Min. 2007. “10. Indices of Robustness for Sample Representation.” *Sociological Methodology* 37 (1): 349–92. <https://doi.org/10.1111/j.1467-9531.2007.00186.x>.
- Frank, K. A., G. Sykes, D. Anagnostopoulos, M. Cannata, L. Chard, A. Krause, and R. McCrory. 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>.
- Frank, Ken, Qinyun Lin, Spiro Maroulis, Shimeng Dai, Nicole Jess, Hung-chang Lin, Sarah Maestrales, Ellen Searle, Jordan Tait, et al. 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).
- Razavi, S., A. Jakeman, A. Saltelli, C. Prieur, B. Iooss, E. Borgonovo, E. Plischke, et al. 2021. “The Future of Sensitivity Analysis: An Essential Discipline for Systems Modeling and Policy Support.” *Environmental Modelling & Software* 137 (March). <https://doi.org/10.1016/j.envsoft.2020.104954>.
- Xu, R., K. A. Frank, S. J. Maroulis, and J. M. Rosenberg. 2019. “Konfound: Command to Quantify Robustness of Causal Inferences.” *The Stata Journal* 19 (3): 523–50. <https://doi.org/10.1177/1536867X19874223>.