

Empirical calibration

Martijn Schuemie

2015-03-31

Contents

1	Introduction	1
2	Negative controls	2
2.1	Plot negative control effect sizes	2
3	Empirical null distribution	3
3.1	Fitting the null distribution	3
3.2	Evaluating the calibration	4
3.3	Plotting the null distribution	4
4	P-value calibration	5
4.1	Calibrating the p-value	5
4.2	Plotting the null distribution	5
5	References	6

1 Introduction

In observational studies, there is always the possibility that an effect size estimate is biased. This can be true even for advanced, well thought out study designs, because of unmeasured or unmodeled confounding. Negative controls (test-hypotheses where the exposure is not believed to cause the outcome) can be used to detect the potential for bias in a study, and with enough negative controls we can start to estimate the systematic error distribution inherent in an observational analysis. We can then use this estimated distribution to compute a calibrated p-value, which reflects the probability of observing an effect size estimate when the null hypothesis (of no effect) is true, taking both systematic and random error into account.

In this document we will use an example study to illustrate how this can be done using the `EmpiricalCalibration` R package. In the example, we will try to answer the question whether sertraline (an SSRI) causes GI bleeding. We use a Self-Controlled Case Series (SCCS) design, and have applied this to a large insurance claims database.

The results from this study are available in the package, and can be loaded using the `data()` command:

```
data(sccs)
drugOfInterest <- sccs[sccs$groundTruth == 1, ]
drugOfInterest
```

```
##      drugName groundTruth      logRr      seLogRr
## 6 Sertraline           1 0.7326235 0.07371708
```

```
exp(drugOfInterest$logRr)
```

```
## [1] 2.080532
```

```
computeTraditionalP(drugOfInterest$logRr, drugOfInterest$seLogRr)
```

```
## [1] 0
```

Here we see that the effect estimate for sertraline is 2.1, with a p-value that is so small R rounds it to 0.

2 Negative controls

Negative controls are drug-outcome pairs where we believe the drug does not cause (or prevent) the outcome. In other words, we believe the true effect size to be a relative risk of 1. We would prefer our negative controls to have some resemblance with our hypothesis of interest (in our example sertraline - GI bleed), and we therefore typically pick negative controls where the outcome is the same (exposure controls), or the exposure is the same (outcome controls). In this example, we have opted for exposure controls, and have identified a set of drugs not believed to cause GI bleed. We have executed exactly the same analysis for these exposures, resulting in a set of effect size estimates, one per negative control:

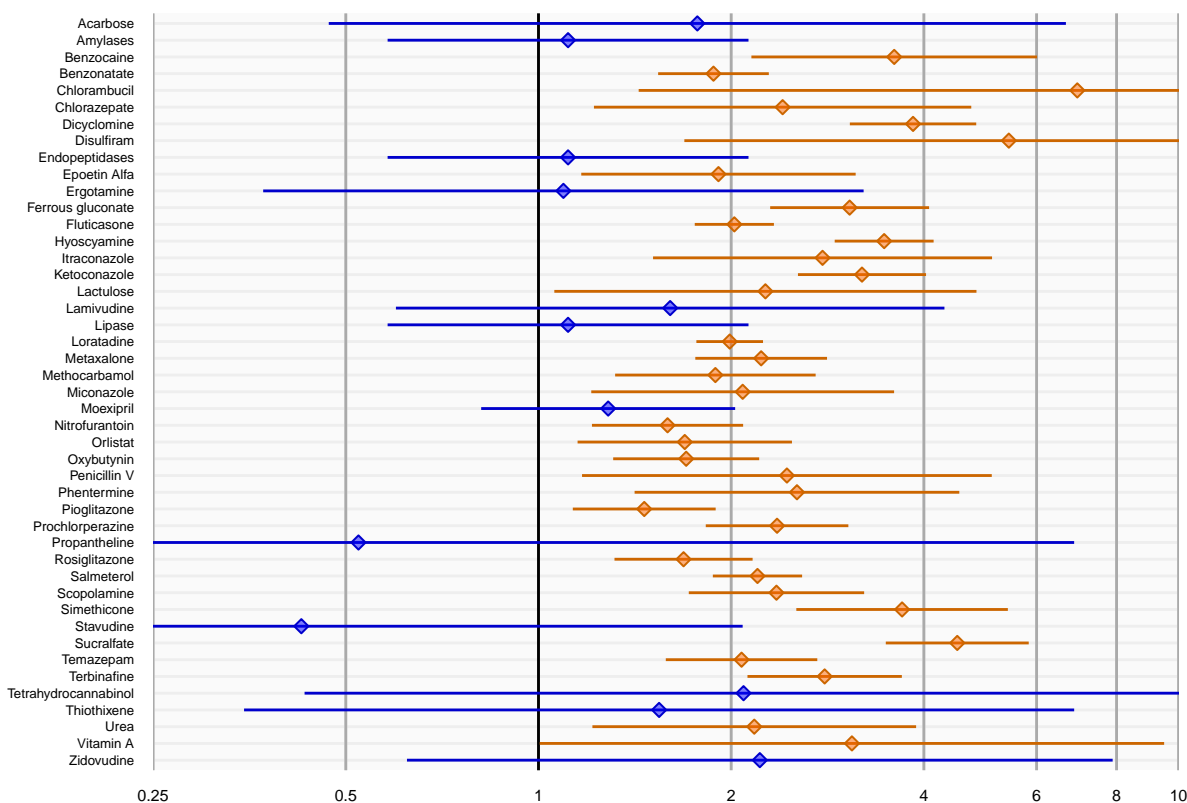
```
data(sccs)
negatives <- sccs[sccs$groundTruth == 0, ]
head(negatives)
```

```
##      drugName groundTruth    logRr    seLogRr
## 1  Thiothixene           0 0.4339021 0.7617538
## 2  Methocarbamol         0 0.6363184 0.1839892
## 4   Phentermine          0 0.9297549 0.2979802
## 5   Disulfiram           0 1.6919273 0.5955222
## 7    Orlistat            0 0.5261691 0.1967199
## 8 Prochlorperazine       0 0.8581890 0.1308460
```

2.1 Plot negative control effect sizes

We can start by creating a forest plot of our negative controls:

```
plotForest(negatives$logRr, negatives$seLogRr, negatives$drugName)
```



Here we see that many negative controls have a confidence interval that does not include a relative risk of 1 (orange lines), certainly more than the expected 5%. This indicates the analysis has systematic error.

3 Empirical null distribution

3.1 Fitting the null distribution

We can use the negative controls to estimate the systematic error distribution. We assume the distribution is a Gaussian distribution, which we have found to give good performance in the past.

```
null <- fitNull(negatives$logRr, negatives$seLogRr)
null
```

```
## Estimated null distribution
##
##      Estimate lower .95 upper .95
## Mean  0.79210    0.68030    0.9039
## SD    0.28340    0.20382    0.3940
```

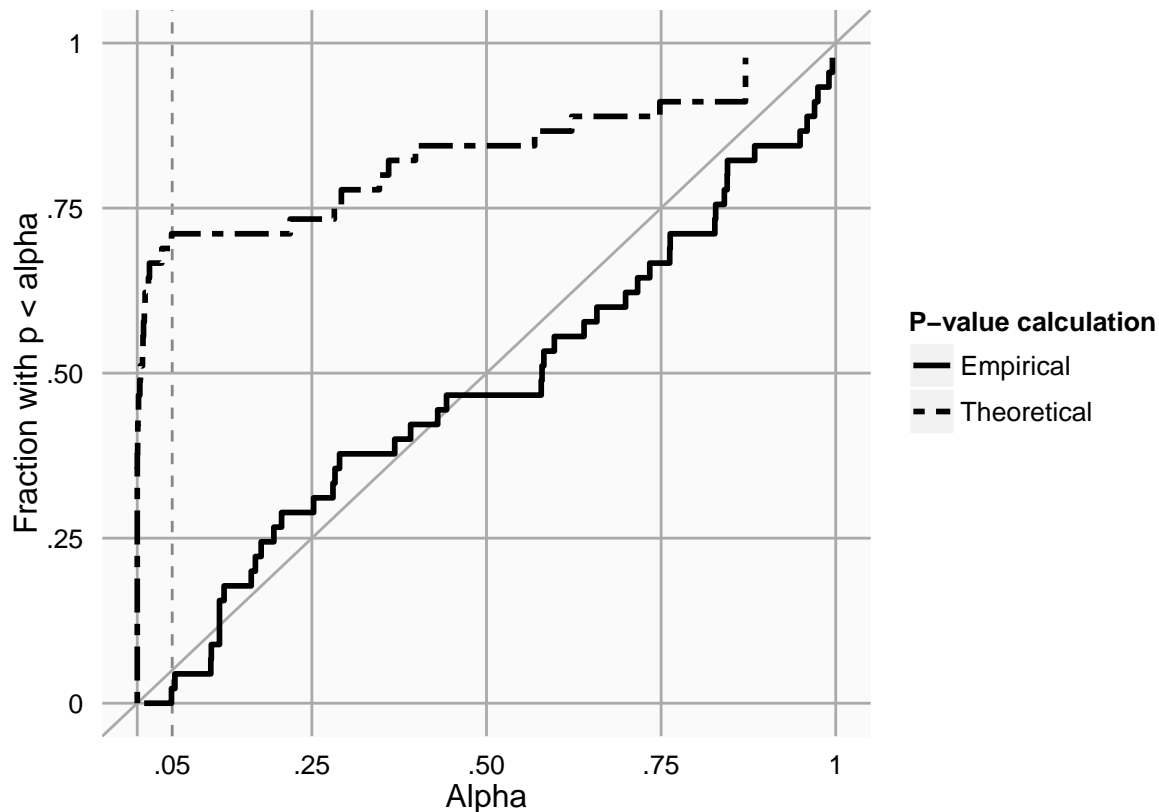
We see that the point estimate of the mean of our distribution is greater than 0, indicating the analysis is positively biased. We also see the point estimate of the standard deviation is greater than 0.25, indicating there is considerable variability in the systematic error from one estimate to the next.

We also see that there is uncertainty around the estimates of the mean and standard deviation, as expressed in the 95% confidence intervals. This uncertainty can be reduced by either increasing the number of negative controls, or by increasing the power for the existing controls (e.g. by waiting for more data to accumulate).

3.2 Evaluating the calibration

To evaluate whether our estimation of the systematic error distribution is a good one, we can test whether the calibrated p-value is truly calibrated, meaning the fraction of negative controls with a p-value below alpha is approximately the same as alpha:

```
plotCalibration(negatives$logRr,negatives$seLogRr)
```



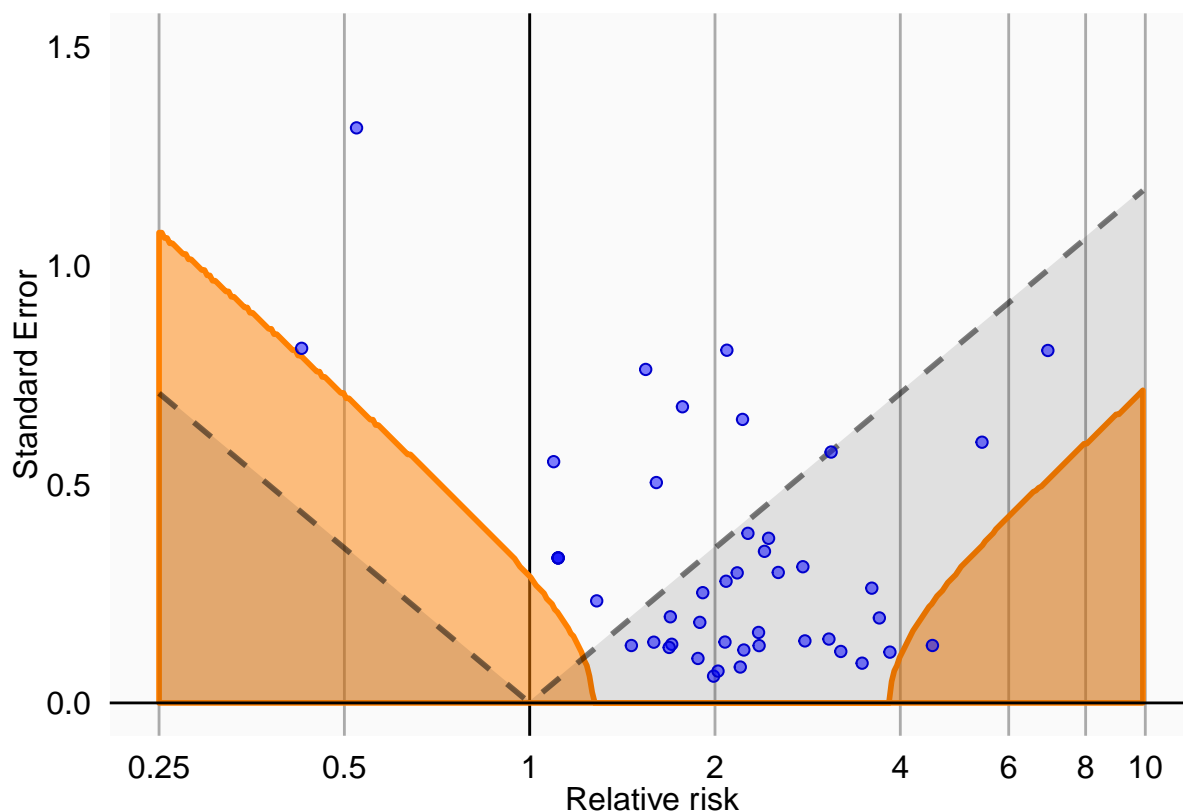
This method uses a leave-one-out design: for every negative control, the null distribution is fitted using all other negative controls, and the calibrated p-value for that negative control is computed.

In the graph we see that the calibrated p-value is much closer to the diagonal than the uncalibrated p-value.

3.3 Plotting the null distribution

We can create a graphical representation of the null distribution, together with the negative controls used to estimate that distribution:

```
plotCalibrationEffect(negatives$logRr,negatives$seLogRr, null = null)
```



In this graph, the blue dots represent the negative controls. Any estimates below the gray dashed lines will have a traditional p-value below .05. In contrast, only estimates that fall within the orange areas will have a calibrated p-value below .05.

4 P-value calibration

4.1 Calibrating the p-value

We can now use the estimated null distribution to compute the calibrated p-value for our drug of interest:

```
p <- calibrateP(drugOfInterest$logRr, drugOfInterest$seLogRr, null, TRUE)
p
```

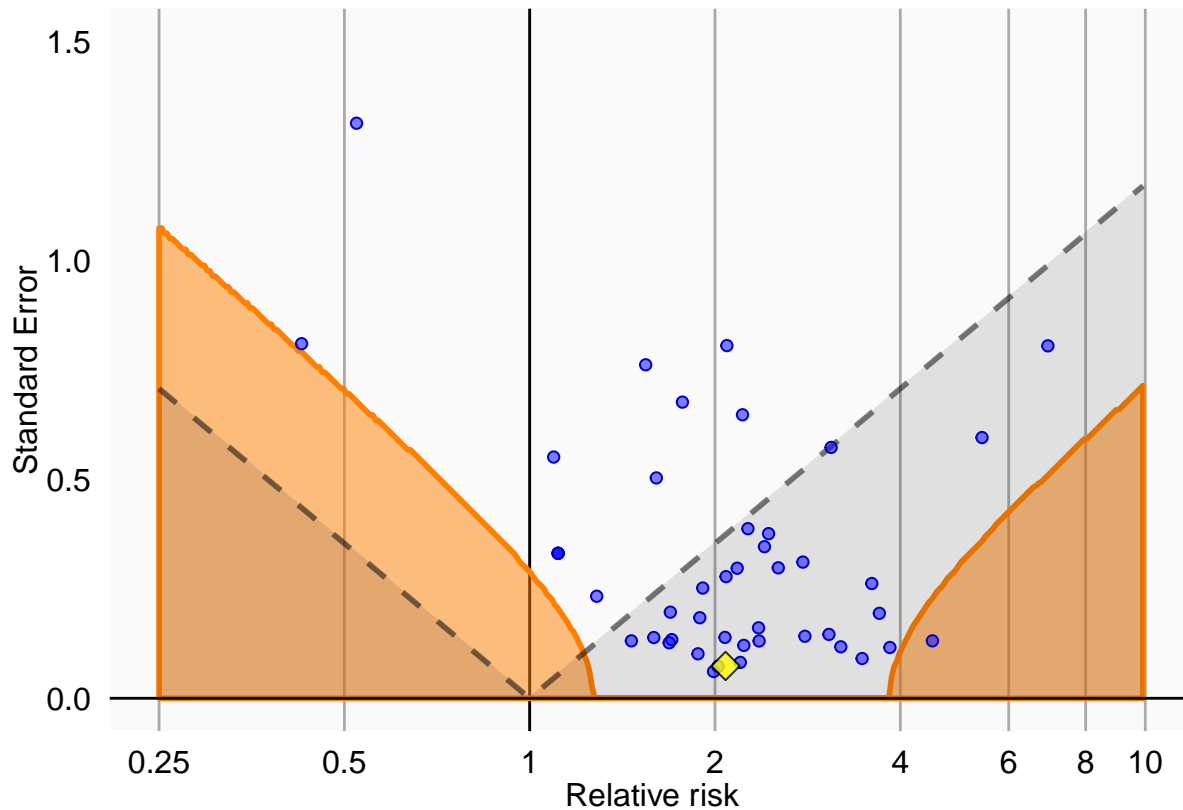
```
##           p    lb95ci ub95ci
## 1 0.8390598 0.5222502 0.99215
```

In this case, the calibrated p-value is 0.84, meaning we have very little confidence we can reject the null hypothesis. Also note that there is uncertainty around the calibrated p-value as expressed in the 95% confidence interval of the p-value. This is caused by the uncertainty in the estimation of the null distribution.

4.2 Plotting the null distribution

A visual representation of the calibration makes it clear why we are no longer certain we can reject the null hypothesis:

```
plotCalibrationEffect(negatives$logRr,
                     negatives$seLogRr,
                     drugOfInterest$logRr,
                     drugOfInterest$seLogRr,
                     null)
```



In this plot we see that, even though the drug of interest (the yellow diamond) has a high relative risk, it is indistinguishable from our negative controls.

5 References

```
citation("EmpiricalCalibration")
```

```
##
## To cite EmpiricalCalibration in publications use:
##
## Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA and Madigan D
## (2014). "Interpreting observational studies: why empirical
## calibration is needed to correct p-values." Statistics in
## Medicine, 33(2), pp. 209-218. <URL:
## http://onlinelibrary.wiley.com/doi/10.1002/sim.5925/abstract>.
##
## A BibTeX entry for LaTeX users is
```

```

##
## @Article{,
##   author = {M. J. Schuemie and P. B. Ryan and W. DuMouchel and M. A. Suchard and D. Madigan},
##   title = {Interpreting observational studies:
##     why empirical calibration is needed to correct p-values},
##   journal = {Statistics in Medicine},
##   volume = {33},
##   number = {2},
##   pages = {209-218},
##   year = {2014},
##   url = {http://onlinelibrary.wiley.com/doi/10.1002/sim.5925/abstract},
## }

```