

Analyzing DECREASE trials for extent of data fabrication

Chris HJ Hartgerink, Gerben ter Riet, Marleen Kemper

29 April, 2016

The effectiveness of beta-blockers in preventing perioperative mortality in non-cardiac surgery has been subject of discussion due to findings of misconduct.¹⁻⁴. Perioperative mortality is the deathrate of patients during the period of the surgical procedure, typically including admission, anaesthesia, surgery, and recovery.

The trials that were found to be suspect were the DECREASE trials¹⁻³. Upon being discovered as (potentially) fabricated, the DECREASE studies were excluded from a 2014 meta-analysis⁴. The meta-analyses restricted to non-DECREASE studies found that conclusions on the effectiveness of beta-blockers in perioperative mortality were reversed — beta-blockers seemed to actually *increase* perioperative mortality instead of decreasing it.

The committee reports about the DECREASE studies indicate that data fabrication was likely¹⁻³, but the extent of the data fabrication remained unclear. In this paper, we aim to estimate the extent of data fabrication in the DECREASE studies^{5,6}.

To investigate the evidence for data fabrication in the DECREASE studies^{5,6}, we took three steps. First, we tried to replicate the findings from the 2014 meta-analysis⁴. Second, we evaluated the veracity of the DECREASE studies (i.e., the probability that these study findings occurred assuming no data fabrication). Third, we reversed this assumption and assumed that data fabrication did occur, estimating how many data points would have to be fabricated to reproduce the results of the DECREASE studies.

Step 1: reproducing Bouri et al. (2014)

Methods

To ensure that we were using similar analytic procedures as in the 2014 meta-analysis⁴, we initially reproduce Bouri et al.’s estimates. This ensured that (1) their results are reproducible and (2) we are using the correct estimates in the following steps. Using figures 2 and 3 from the original paper⁴, we extracted the raw event data for the 2 (control vs experimental) by 2 (event or no event) design, which we used to recompute the natural logarithm of the risk ratio and its standard error. The extracted event data is available at osf.io/aykeh and our analysis plan was preregistered at osf.io/vnmzc.

We computed the log risk ratio (i.e., log RR) for each study and pooled these using the R package `metafor`⁷. We estimated a weighted random-effects model using the restricted maximum-likelihood estimator (i.e., REML)⁸ to estimate the variance of effects. We used the default weighting procedure in the `metafor` package. When there was a zero-count for a cell, .5 was added to each cell of that trial⁹. The 2014 meta-analysis⁴ did not specify the effect variance estimate used. Hence, minor discrepancies between our estimates and the original estimates might be due to differences in the estimation procedure.

Results

First, we analyzed the RRs to verify the estimates in Figure 2 and found that we were able to reproduce the estimates for the different sets of studies. The original Figure 2⁴ differentiated between the estimates from the non-DECREASE trials and the DECREASE trials. We confirmed the estimates of the effect size and the variance for both the non-DECREASE- and the DECREASE trials, save for some minor discrepancies due to the estimation method. Table 1 depicts the original and reproduced values for both sets of studies.

Table 1. The original- and reproduced meta-analytic results based on the data provided in the 2014 meta-analysis.⁴

		Risk ratio	τ^2	Confidence interval
Non-DECREASE	Original	1.27	0.00	[1.01; 1.60]
	Reproduced	1.28	0	[1.01; 1.62]
DECREASE	Original	0.42	0.29	[0.15; 1.23]
	Reproduced	0.44	0.24	[0.15; 1.23]

Second, we meta-analyzed all studies combined, including a dummy predictor for the DECREASE and non-DECREASE studies to replicate results presented in Figure 4 of the 2014 meta-analysis.⁴ Surprisingly, our results showed more evidence against equal subgroups than the original meta-analysis⁴. Whereas the original test for subgroup differences was $\chi^2(1) = 3.91, p = .05$, our analyses indicated that $\chi^2(1) = 6.12, p = 0.013$. Additionally, the original analyses showed substantial heterogeneity ($I^2 = 74.4\%$) where we found an $I^2 = 0\%$. Different estimates of the heterogeneity (e.g., DerSimonian-Laird instead of REML) did not change this difference.

Step 2: evaluating the veracity of DECREASE studies

We estimated the probability of obtaining the results in the DECREASE studies based on the estimates from the non-DECREASE studies. To this end, we assumed that the non-DECREASE studies provide a valid representation of the true effect of beta-blockers on perioperative mortality. The estimated probability is also known as the veracity of the data¹⁰, which indicates the probability of the observed data under a given true effect. We assumed that the non-DECREASE studies estimate the true population distribution of effects, not perturbed by publication bias due to statistical (non)significance. Publication bias was assumed to not be a problem because a substantial number of nonsignificant effects are included in the dataset.

Method

Based on the estimated mean log RR and credible interval in the non-DECREASE studies, we computed the probability of log RR in the DECREASE trials. The estimates of the non-DECREASE studies were obtained from step 1, which include the estimated log RR (i.e., 0.25), and the 95% credibility interval as provided by the package `metafor` (i.e., [0.01; 0.484]). We assumed a normal distribution of population effects with the estimated effect as the mean of the distribution. The 95% credibility interval denotes the bounds of the normal distribution that covers 95% of the density, where the standard deviation is calculated as the distance from the mean to either bound, divided by 1.96. This allows for an approximation of the population effect distribution, as depicted in Figure 1.

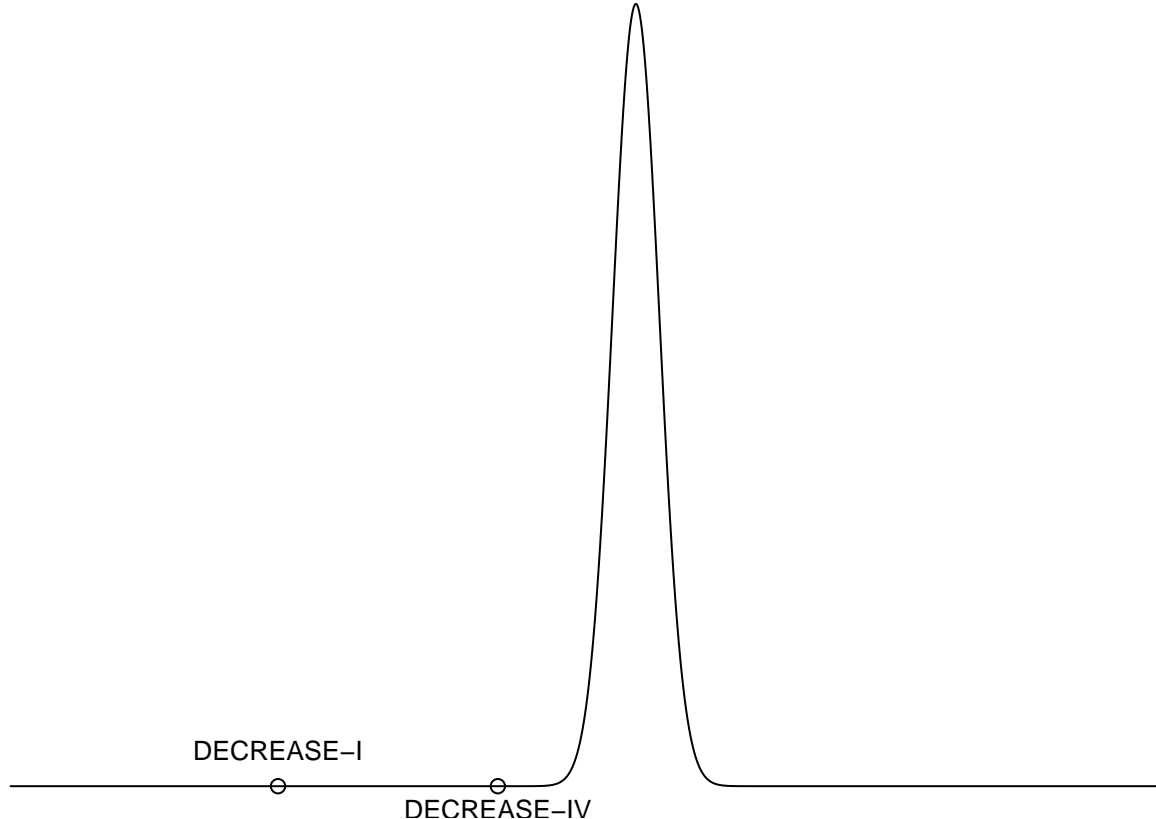


Figure 1. Density plot of the estimated true effect distribution based on the non-DECREASE studies, with the position of the DECREASE studies highlighted.

Based on the estimated effect distribution from the non-DECREASE trials, we calculated the probability of each DECREASE trial result, or a more extreme result. In other words, we computed the p -value for the hypothesis that the DECREASE studies occurred based on the information available from the other trials.

Results

Figure 1 indicates that the DECREASE trials are highly unlikely under the estimated effect distribution based on the non-DECREASE trials. More specifically, DECREASE-I (or more extreme) has a probability of $2.9610213 \times 10^{-53}$ (less than 1 in a quintillion) and DECREASE-IV 3.0974376×10^{-9} (3 in a billion), which indicates they are unlikely to come from the same population effect distribution. Observing two of such extremely unlikely results in this population effect distribution is highly improbable, $9.1715786 \times 10^{-62}$. This would therefore indicate that the DECREASE trials are severely different from the non-DECREASE studies.

We note that the variance in the non-DECREASE studies was estimated at 0, indicating that there would supposedly be a fixed-effect of beta-blockers on perioperative mortality despite differences in type of beta-blockers administered and dosage/duration of the beta-blockers. This homogeneity is not excessive, considering that the Q -statistic is sufficiently large to begin with (i.e., 5).¹¹ Nonetheless, the original non-DECREASE trials show substantive variation in the treatment, for example the usage of different beta-blockers (i.e., atenolol, propranolol, bisoprolol, metoprolol) in different (standardized) dosages, which have been argued to affect effectiveness of the interventions. As such, the lack of heterogeneity can be genuine and indicative of no difference between these different beta-blockers and their dosage/duration. Nonetheless, it could also be due to uncertainty in the estimate of heterogeneity τ^2 due to the small set of studies included.

Considering that the variance (i.e., heterogeneity) estimate is 0 and uncertain due to the small number of studies included, sensitivity analyses of the results are in place. The probability of observing both DECREASE trials stays approximately below 1 out of 1000 until the variance estimate is .25, and the probability stays

approximately below 1 out of 100 until the variance estimate is .43 (see Figure 2). Further research would be worthwhile to better estimate the variance in the effects of beta-blockers on perioperative mortality.

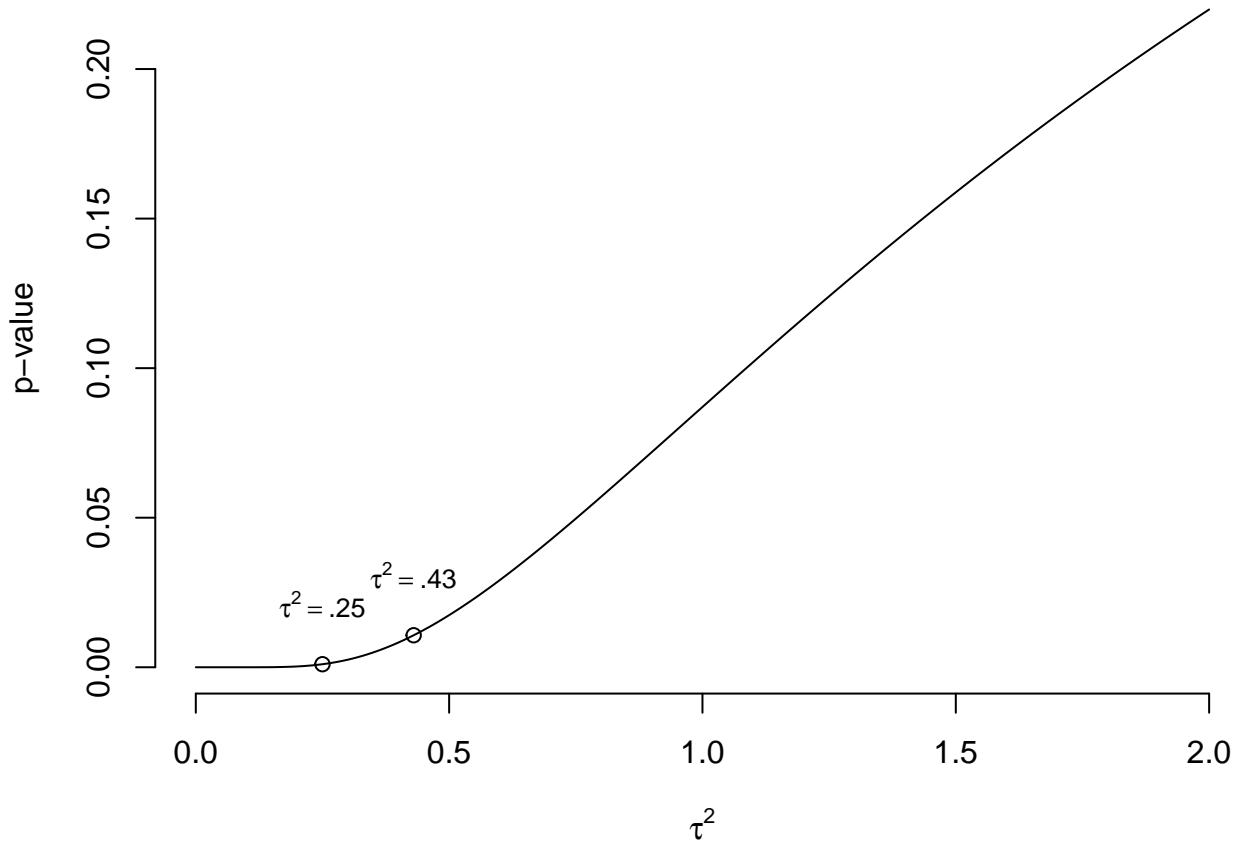


Figure 2. Sensitivity analyses for the p -value that indicates the probability of observing the results from the DECREASE studies, or more extreme results, based on the estimated true effect (non-DECREASE trials) and the accompanying variance estimate.

Step 3: estimating the amount of fabricated data

We estimated the number of data points that would need to be fabricated to arrive at the estimates from the DECREASE studies, given that the non-DECREASE studies represent the true effect. This assumes, in contrast to Step 2, that the DECREASE studies in fact contain fabrication data. The estimates from Step 3 provide an indication of the extent of the potential data fabrication in the DECREASE studies under this assumption.¹⁻⁴

Method

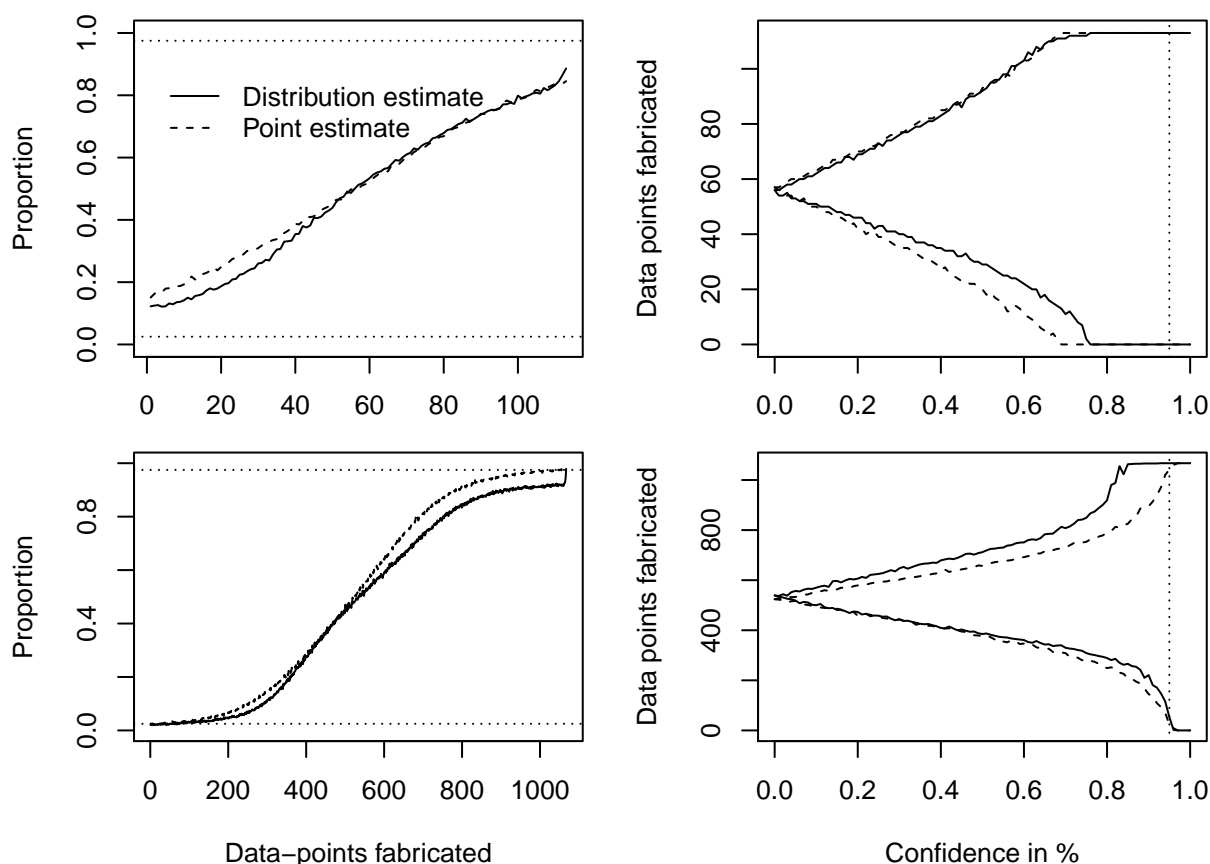
In order to estimate the number of fabricated data points, we meta-analyzed the log odds of mortality per condition for both the DECREASE- and non-DECREASE studies separately. For each of these, we ran a meta-analysis applying the same methods as in Step 1. These separate analyses were necessary, because Step 1 only provided estimates of the overall risk ratio, but not of the odds conditional on exposure (beta-blocker or control) and study set (DECREASE- versus non-DECREASE).

We applied the inversion method to estimate the number of data points fabricated in the DECREASE-I and DECREASE-IV studies.¹² The inversion method iteratively simulates data with X fabricated data points; X varying from 0 through N . For each X , 10,000 iterations were conducted. Data fabrication occurred using a

binomial distribution, with the population effect from the DECREASE studies serving as the probability of an event occurring. The population effect was either drawn from the estimated population distribution, which introduced uncertainty (i.e., uncertain estimate), or was equated to the point estimate (i.e., certain estimate; not preregistered). The remaining genuine data, $N - X$, were simulated in a similar fashion, where we only used the effect distribution from the non-DECREASE studies. Using the (partially) fabricated dataset, we estimated the risk ratio from the data and noted whether the two-sided p -value was larger under the non-DECREASE effect distribution (coded 0) or the DECREASE effect distribution.

Based on the results of the inversion method, we obtain a confidence interval of data points fabricated in the DECREASE studies. After running all iterations for one X , the proportion of trials that were more likely under the DECREASE population effect than under the non-DECREASE population effect were computed. This resulted in a proportion for each X from 0 to N . Those proportions closest to .025 and .975 denoted the 95% confidence interval of data points fabricated.

Results



Discussion

References

1. Onderzoekscommissie Wetenschappelijke Integriteit. *Onderzoek Naar Mogelijke Schending van de Wetenschappelijke Integriteit: Beknopte Versie*. Erasmus MC; 2011. <https://web.archive.org/web/20151113121125/>

<http://www.erasmusmc.nl/cs-research/bijlagen/integriteit/rapport-poldermans-2011>.

2. Commissie Vervolgonderzoek 2012. *Rapport Vervolgonderzoek Naar Mogelijke Schending van de Wetenschappelijke Integriteit*. Erasmus MC; 2012. <https://web.archive.org/web/20151027084205/http://www.erasmusmc.nl/5663/135857/3675250/3706798/erasmusmc.commissie.verv.onderzoek.2012>.
3. Commissie Vervolgonderzoek Wetenschappelijke Integriteit 2013. *Rapport*. Erasmus MC; 2014.
4. Bouri S, Shun-Shin MJ, Cole GD, Mayet J, Francis DP. Meta-analysis of secure randomised controlled trials of β -blockade to prevent perioperative death in non-cardiac surgery. *Heart*. 2014;100(6):456-464. doi:<http://doi.org/10.1136/heartjnl-2013-304262>.
5. Dunkelgrun M, Boersma E, Schouten O, et al. Bisoprolol and fluvastatin for the reduction of perioperative cardiac mortality and myocardial infarction in intermediate-risk patients undergoing noncardiovascular surgery: A randomized controlled trial (DECREASE-IV). *Annals of surgery*. 2009;249(6):921-926. doi:<http://doi.org/10.1097/SLA.0b013e3181a77d00>.
6. Poldermans D, Boersma E, Bax JJ, et al. The effect of bisoprolol on perioperative mortality and myocardial infarction in High-Risk patients undergoing vascular surgery. *The New England journal of medicine*. 1999;341(24):1789-1794. doi:<http://doi.org/10.1056/NEJM199912093412402>.
7. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*. 2010;36:1-48. <http://www.jstatsoft.org/v36/i03/>.
8. Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the Random-Effects model. *Journal of educational and behavioral statistics: a quarterly publication sponsored by the American Educational Research Association and the American Statistical Association*. 2005;30(3):261-293. doi:<http://doi.org/10.3102/10769986030003261>.
9. Agresti A. *Categorical Data Analysis*. Hoboken, NJ: John Wiley & Sons Inc.; 2002.
10. Peters CFW, Klaassen CAJ, Wiel MA van de. *Evaluating the Scientific Veracity of Publications by Dr. Jens Forster*. University of Amsterdam; 2015.
11. Ioannidis JPA, Trikalinos TA, Zintzaras E. Extreme between-study homogeneity in meta-analyses could offer useful insights. *Journal of clinical epidemiology*. 2006;59(10):1023-1032. doi:<http://doi.org/10.1016/j.jclinepi.2006.02.013>.
12. Casella G, Berger RL. *Statistical Inference*. Pacific Grove, CA: Duxbury; 2002.