

# Analyzing DECREASE trials for extent of data fabrication

*Chris HJ Hartgerink, Gerben ter Riet, Marleen Kemper*

*01 February, 2017*

The effect of beta-blockers on perioperative mortality in non-cardiac surgery for patients with coronary artery disease (CAD) has been subject of discussion<sup>1</sup> due to findings of research misconduct<sup>2-4</sup> that were central to claims of their effectiveness<sup>5-7</sup>. Perioperative mortality is the deathrate of patients during the period of the surgical procedure, which typically includes admission, anaesthesia, surgery, and recovery. This related finding of research misconduct altered meta-analytic results of perioperative mortality in non-cardiac surgery for CAD patients substantively<sup>7</sup>, where meta-analyses that include these studies conclude that beta-blockers decrease perioperative mortality<sup>5-7</sup> but a meta-analysis that excludes these studies concludes that beta-blockers increase perioperative mortality<sup>7</sup>.

The trials deemed to be the result of research misconduct were the Dutch DECREASE trials<sup>2-4</sup>. The committees that investigated the integrity of the DECREASE studies reported that data fabrication was likely<sup>2-4</sup>, but that the extent of the data fabrication remained unclear. Given that the inclusion or exclusion of these trials in meta-analyses is crucial in determining whether beta-blockers are indeed effective during the perioperative period and the latest guidelines still recommend the usage of beta-blockers<sup>8</sup>, we aim to estimate the extent of data fabrication in the DECREASE studies<sup>9,10</sup> in order to further this debate.

The reports on the DECREASE trials primarily focused on the provenance of the raw data and patient files (e.g., informed consent, whether data corresponded to patient files), but neglected the extent to which the DECREASE studies deviated from similar trials that tested the effectiveness of beta-blockers on perioperative mortality. Comparing how (dis)similar trials are from comparable studies is not a new idea<sup>11</sup> and has proven effective in detecting data fabrication<sup>12</sup>. This is a different approach from the forensic statistical methods previously applied<sup>4</sup>, where results within one study are evaluated as opposed to other similar studies. This explains the discrepancy between the feasibility of evaluating the DECREASE trials by the statistical expert of the committee report and ours.

To statistically investigate the evidence of data fabrication in the results of the DECREASE studies<sup>9,10</sup>, we took three steps. First, we replicate the findings from a 2014 meta-analysis<sup>7</sup> that contains sufficient information to estimate the deviation of the DECREASE-I and DECREASE-IV studies from the remaining studies. Second, we evaluate the veracity of the DECREASE-I and DECREASE-IV studies (i.e., the probability that these study findings occurred assuming no data fabrication occurred). Third, we reverse this assumption and assume that data fabrication did occur and estimate how many data points would have to be fabricated to reproduce the results of the DECREASE-I and DECREASE-IV studies if the remaining studies are regarded as estimating the true effect of beta-blockers on perioperative mortality.

## Step 1: reproducing meta-analysis of Bouri et al. (2014)

### Methods

To ensure that we used similar analysis procedures as in the 2014 meta-analysis<sup>7</sup>, we initially reproduced Bouri et al.'s estimates. This ensured that (1) their results are reproducible and (2) we are using the correct estimates in subsequent steps of our analyses. Using figures 2 and 3 from the original paper<sup>7</sup>, 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](https://osf.io/aykeh) and our analysis plan was preregistered at [osf.io/vnmzc](https://osf.io/vnmzc).

We computed the log risk ratio (i.e., log RR) for each study and pooled these using the R package `metafor`<sup>13</sup>. We estimated a weighted random-effects model using the restricted maximum-likelihood estimator (i.e., `REML`)<sup>14</sup> 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 trial, as is common in meta-analyses on risk- and odds ratios in order to deal with computational artefacts<sup>15</sup>. The 2014 meta-analysis<sup>7</sup> did not specify the effect variance estimate used; hence, minor discrepancies between our estimates and the original estimates could be due to differences in the estimation procedure.

## Results

We analyzed the RRs to verify the original estimates (Figure 2 of the 2014 meta-analysis<sup>7</sup>) and found that we were able to reproduce the estimates for the different sets of studies. Bouri et al. differentiated between the estimates from the non-DECREASE trials and the DECREASE trials. We confirmed the effect size estimates and the variance estimates 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 by Bouri et al.

|              |            | Risk ratio | $\tau^2$ | Confidence interval |
|--------------|------------|------------|----------|---------------------|
| Non-DECREASE | Original   | 1.27       | 0        | [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<sup>7</sup>. Surprisingly, our results showed more evidence for unequal subgroups than the original meta-analysis<sup>7</sup>; 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 heterogeneity estimators (e.g., `DerSimonian-Laird` instead of `REML`) did not resolve this difference. We tried to clarify these discrepancies by e-mailing the original authors, but failed to receive a response.

## Step 2: evaluating the veracity of DECREASE studies

Based on the non-DECREASE estimates from Step 1, we estimated the probability of obtaining the results in the 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 (similar to Bouri et al.<sup>7</sup>; but see the discussion). The estimated probability is also known as the veracity of the data<sup>16</sup>, 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 (9 of 11 nonsignificant).

## Method

Based on the estimated mean log RR and its 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 its 95% credibility interval as

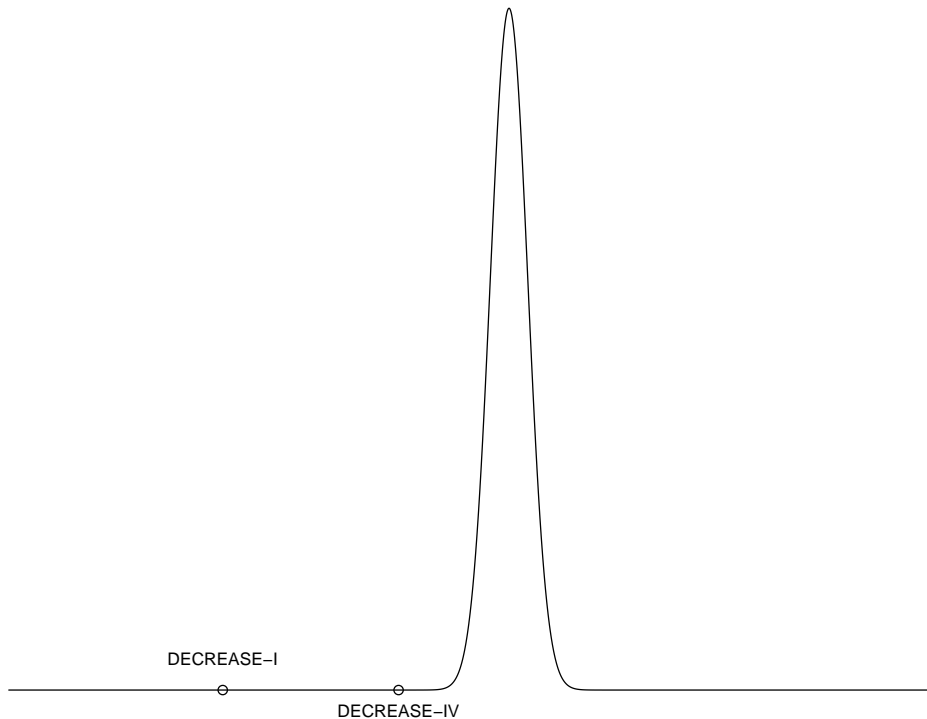


Figure 1: Density plot of the estimated true effect distribution based on the non-DECREASE studies only, with the position of the DECREASE studies highlighted.

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.

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  $3.0974376 \times 10^{-9}$  (3 in a billion), which indicates they are unlikely to come from the same population effect distribution as the non-DECREASE trials. Observing two of such extremely unlikely results in this population effect distribution jointly is highly improbable,  $9.1715786 \times 10^{-62}$ . This would therefore indicate that the DECREASE trials are severely different from the non-DECREASE studies.

Results from Step 1 indicated that no variance (i.e., homogeneity;  $\tau^2 = 0$ ) of the effects was observed; given the small number of studies included (i.e., 9) this estimate is uncertain, however. We conducted sensitivity analyses in order to see how dependent results are on the specific heterogeneity estimate. 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

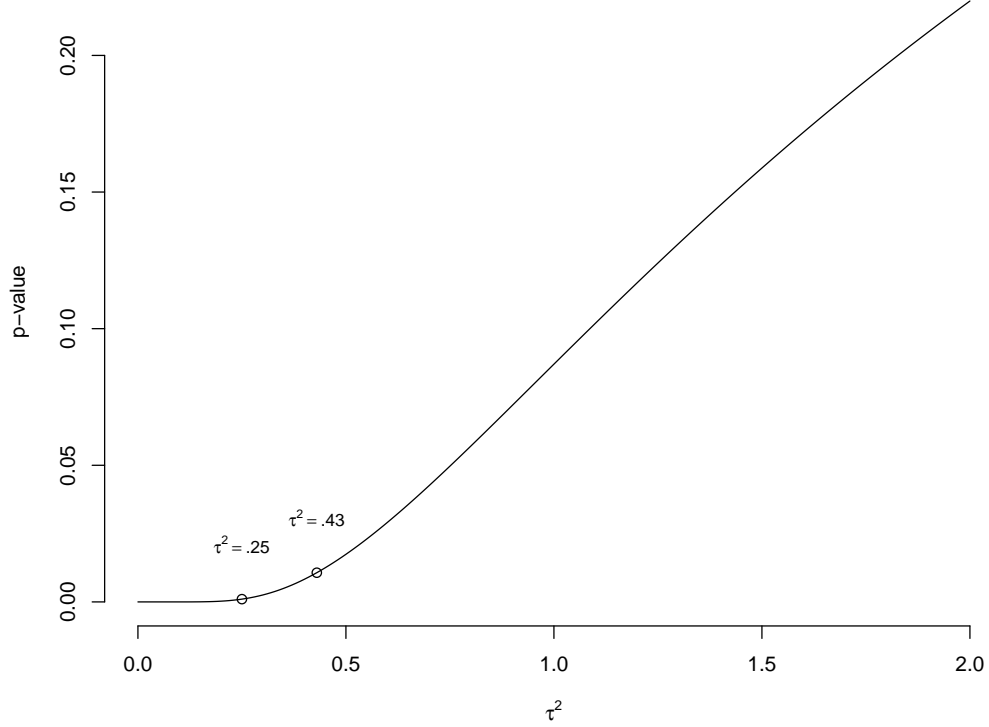


Figure 2: Sensitivity analyses for the p-value that indicates the probability of observing the results from the DECREASE studies, or more extreme results, based on the estimated true effect (non-DECREASE trials) and the accompanying variance estimate.

perioperative mortality and what moderators affect the effect size (e.g., type of beta-blocker).

### 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 fabricated data. The estimates from Step 3 provide an indication of the extent of the potential data fabrication in the DECREASE studies under this assumption<sup>2-4,7</sup>.

#### Method

In order to estimate the number of fabricated data points, we estimated the population effect distribution of mortality (in log odds) per condition in both the DECREASE- and non-DECREASE studies separately. For each of these, we ran a meta-analysis applying the same, verified methods used in Step 1. These separate analyses were necessary, considering that Step 1 only provided estimates of the overall risk ratio but not the odds conditional on exposure (beta-blocker or control) and study set (DECREASE- versus non-DECREASE). This resulted in four meta-analytic mortality estimates with an accompanying variance. Throughout the simulations, we use the point estimates to simulate genuine- and fabricated data, but supplement this by using the distribution estimates as sensitivity analyses.

We applied the inversion method to estimate the number of fabricated data points in the DECREASE-I and DECREASE-IV studies<sup>17</sup>. With the inversion method, we iteratively hypothesized that  $X$  out of  $N$  data

points were fabricated (i.e.,  $0, 1, \dots, N$ ) in the DECREASE-I and DECREASE-IV studies separately. For each study, we simulated 10,000 datasets per hypothesis of  $X$  fabricated data points and  $N-X$  genuine data points. For each simulated dataset (exact simulation procedure in next paragraph), we determined the likelihood of the results with

$$L(\theta|\pi_E, \pi_C) = \pi_E^{n_{11}}(1 - \pi_E)^{n_{12}} \times \pi_C^{n_{21}}(1 - \pi_C)^{n_{22}} \quad (1)$$

where  $\pi_E$  indicates the mortality rate of the experimental condition as drawn from the meta-analytic effect distribution ( $\pi_C$  indicates the same for the control condition). The likelihood was computed under both the fabricated effect estimates (i.e.,  $L_{fabricated}$ ) and the genuine data (i.e.,  $L_{genuine}$ ). Table 2 indicates which cell sizes the various  $n_{XX}$  refer to within the (simulated) data. We compared the likelihoods to determine whether the simulated data were more likely to arise from the genuine- ( $L_{genuine}$ ) or from the fabricated data ( $L_{fabricated}$ ). Note that using the likelihoods for this is a minor deviation from the preregistration, where we initially planned on using  $p$ -value comparisons (osf.io/vnmztc).

Table 2: Outcome possibilities within a simulated 2 (beta-blocker v control) by 2 (dead v alive) clinical trial.

|               | Dead     | Alive    |
|---------------|----------|----------|
| Beta-blockers | $n_{11}$ | $n_{12}$ |
| Control       | $n_{21}$ | $n_{22}$ |

For each hypothesis of  $X$  out of  $N$  fabricated datapoints, we computed the probability that the fabricated data are more likely than the genuine data ( $p_F = P(L_{fabricated} > L_{genuine})$ ). Based on  $p_F$ , we computed the confidence interval for  $X$  (i.e.,  $X_{LB}; X_{UB}$ ). For a 95% confidence interval, the lowerbound is equal to the  $p_F$  closest to .025, whereas the upperbound is equal to the  $p_F$  closest to .975.

We computed  $p_F$  for all possibilities of  $X$  out of  $N$  fabricated datapoints in 10,000 randomly generated datasets, which were generated in three steps. For each dataset we:

1. Sampled (without replacement)  $X$  fictitious participants that would be the result of data fabrication (sampled across all conditions).
2. Determined the population mortality rate for each condition (i.e., for each cell as in Table 3). The meta-analytic point estimate was used or a population effect was randomly drawn from the meta-analytic effect distribution.
3. Simulated the number of deaths for the different conditions using a binomial distribution based on the mortality rate as determined in 2, resulting in the cell counts as in Table 3.

Based on the meta-analytic effect from 2 and the cell sizes from 3, we computed the likelihoods of  $L_{fabricated}$  and  $L_{genuine}$  using Equation 1. As mentioned before,  $p_F$  was computed as the probability that the data were more likely under the estimates resulting from the (allegedly) fabricated data (i.e., the DECREASE trials) than under the estimates resulting from the genuine data (i.e., the non-DECREASE trials;  $p_F = P(L_{fabricated} > L_{genuine})$ ).

## Results

For DECREASE-I, the 95% confidence interval for the estimated number of fabricated data points is [0 - 112] or [3 - 112] when based on a point estimate or a more uncertain distribution estimate, respectively. The left column of Figure 3 depicts the  $p_F$  per  $X$  fabricated data points (top panel) and the bounds of the confidence interval when the degree of confidence is altered (lower panel). Staying clearly between the dotted lines in the top panel, depicting the 95% CI (top: .975; bottom: .025), it becomes apparent that the degree of uncertainty is too high to make any estimates with a high degree of confidence. This due to the small sample size (i.e.,

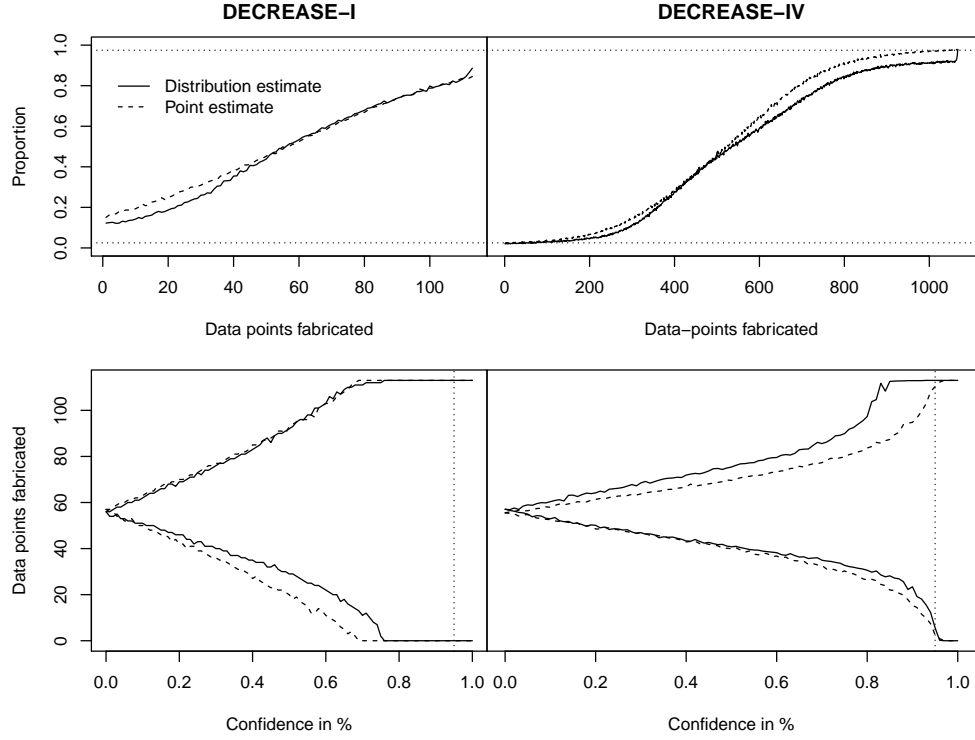


Figure 3: Inversion method results used to estimate the number of data points fabricated in the DECREASE-I and DECREASE-IV trials. The top row panels indicate  $p_F$  (y-axis) for all  $X$  out of  $N$  fabricated data points (x-axis). The bottom row indicates the estimated number of fabricated data points (y-axis) when varying the degree of confidence (x-axis). Dotted lines indicate the bounds for a 95 percent CI.

$N = 112$ ) combined with the availability of just the summary results. Only when the degree of confidence is lowered to around 75% does the interval not span the entire sample size. As such, based on the summary results, little can be said about the extent of the data fabrication that occurred in the DECREASE-I trial, affirming the conclusions of the original committee report<sup>4</sup>.

For DECREASE-IV, the 95% confidence interval for the estimated number of fabricated data points is [8 - 1059] or [7 - 1066] when based on a point estimate or a more uncertain distribution estimate, respectively. The minor difference between the estimates indicates that there is a high degree of confidence that data fabrication did occur. Nonetheless, the range of potentially fabricated data points is still estimated at approximately 1000; this indicates that the summary results are insufficient to provide more than an estimated lowerbound. However, this does indicate that it is possible not all data were fabricated (i.e.,  $N = 1066$ ), increasing the importance of data provenance to discern between genuine and falsified data. This affirms the conclusion of the scientific integrity committee that the results of the DECREASE-IV trial are “scientifically incorrect” (p.11)<sup>3</sup>.

The results also highlight that, despite the lack of availability of raw data, summary results from larger samples result in more certainty about the estimated number of fabricated data points using the inversion method. This is due to decreased standard errors at the population level of the effects, providing less variation in the simulations and resulting in higher sensitivity. However, this also highlights that in order to prevent detection it would be in the fabricators’ interest to fabricate small studies (assuming the fabricator wants to remain undetected).

## Discussion

Throughout steps 1, 2, and 3 we provided evidence that the , the DECREASE trials are highly unlikely to come from a genuine data pool, and that there is a high degree of confidence that data was fab

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).<sup>18</sup> Nonetheless, the original non-DECREASE trials show substantive variation in the treatment, for example the usage of different beta-blockers (i.e., atenolol, propanolol, bisoprolol, metoprolol) in different (standardized) dosages, which have been argued to affect effectiveness of the interventions<sup>19</sup>. 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  $\tau^2$  due to the small set of studies included.

However, due to various types of beta-blockers, evidence might be moderated<sup>19</sup>.

Additionally, during the process of this research paper, the first author tried to ascertain a raw dataset of one of the DECREASE trials via a Freedom of Information request (FOI). The Erasmus MC refused to share these data

The ESC/ESA guidelines from 2014[[@](#)] were published in October 2014, whereas the meta-analysis indicating a reversal of the effectiveness of beta-blockers was published online at the end of July 2013[[@](#)]. The committee revising the ESC/ESA guidelines stated: “The respective writing committees independently performed their literature review and analysis, and then developed their recommendations.”<sup>20</sup> Regardless, the revisions made “recommend continuation of beta-blocker therapy in the perioperative period in patients currently receiving this medication” and do not recommend the “initiation of beta-blockers in patients undergoing low-risk surgery”<sup>21</sup>. Review pieces note no substantial clinical change with respect to beta-blockers.<sup>22,23</sup>

The revisions to the ESC/ESA guidelines do not discourage use of beta-blockers, despite the obvious lack of evidence for systematic evidence of the effectiveness of beta-blockers to reduce perioperative mortality in trials. If anything, the trials show that insufficient evidence has been collected. This becomes apparent when the estimated effect for all non-DECREASE trials is inspected; its  $p$ -value is .04, which is insufficient evidence and is actually more likely when there is no effect in of beta-blockers in reality<sup>24</sup>

In order, the first author tried to ascertain the data of DECREASE VI, the only DECREASE study for which a dataset is still available<sup>1</sup> via a Freedom of Information (FOI) request. Within this paper, we inspected data anomalies based on summary results, which contain less information to be able to determine what is The board of Erasmus Medical Centrum (EMC) was unwilling to comply with this request, stating that it was not a matter of the board. Regardless of whether it was a board related matter,

## References

1. Cole GD, Francis DP. Perioperative beta blockade: Guidelines do not reflect the problems with the evidence from the decrease trials. *BMJ*. 2014;349. doi:10.1136/bmj.g5210.
2. 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>.
3. 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>.
4. Commissie Vervolgonderzoek Wetenschappelijke Integriteit 2013. *Rapport*. Erasmus MC; 2014. <http://web.archive.org/web/20161104135848/http://www.erasmusmc.nl/cs-research/bijlagen/integriteit/eindrapport2014nl>.
5. Devereaux PJ, Beattie WS, Choi PT-L, et al. How strong is the evidence for the use of perioperative beta blockers in non-cardiac surgery? Systematic review and meta-analysis of randomised controlled trials. *BMJ*. 2005;331(7512):313-321. doi:10.1136/bmj.38503.623646.8F.
6. Angeli F, Verdecchia P, Karthikeyan G, Mazzotta G, Gentile G, Reboldi G.  $\beta$ -blockers reduce mortality in patients undergoing high-risk non-cardiac surgery. *American Journal of Cardiovascular Drugs*. 2010;10(4):247-259. doi:10.2165/11539510-000000000-00000.
7. Bouri S, Shun-Shin MJ, Cole GD, Mayet J, Francis DP. Meta-analysis of secure randomised controlled trials of  $\beta$ -blockade to prevent perioperative death in non-cardiac surgery. *Heart*. 2014;100(6):456-464. doi:10.1136/heartjnl-2013-304262.
8. ESC/ESA. Guidelines on non-cardiac surgery: Cardiovascular assessment and management. *European Heart Journal*. 2014;35:2383-2243.
9. 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:10.1097/SLA.0b013e3181a77d00.
10. 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:10.1056/NEJM199912093412402.
11. Buyse M, George SL, Evans S, et al. The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statistics in medicine*. 1999;18(24):3435-3451. doi:10.1002/(SICI)1097-0258(19991230)18:24<3435::AID-SIM365>3.0.CO;2-O.
12. Knepper D, Lindblad AS, Sharma G, et al. Statistical monitoring in clinical trials: Best practices for detecting data anomalies suggestive of fabrication or misconduct. *Therapeutic Innovation & Regulatory Science*. 4-feb 2016. doi:10.1177/2168479016630576.
13. 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/>.
14. 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:10.3102/10769986030003261.

15. Agresti A. *Categorical Data Analysis*. Hoboken, NJ: John Wiley & Sons Inc.; 2002.

16. Peters CFW, Klaassen CAJ, Wiel MA van de. *Evaluating the Scientific Veracity of Publications by Dr. Jens Forster*. University of Amsterdam; 2015.

17. Casella G, Berger RL. *Statistical Inference*. Pacific Grove, CA: Duxbury; 2002.

18. 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:10.1016/j.jclinepi.2006.02.013.

19. Klei W van. Welke perioperatieve bètablokker heeft de voorkeur? [Which perioperative beta-blocker is preferred?]. *Nederlands Tijdschrift voor Geneeskunde*. 2015;159:A9798.

20. Kristensen SD, Knuuti J, Saraste A, et al. 2014 ESC/ESA guidelines on non-cardiac surgery. *European Journal of Anaesthesiology*. 2014;31(10):517-573. doi:10.1097/eja.0000000000000150.

21. Guarracino F PH Baldassarri R. Revised esc/esa guidelines on non-cardiac surgery: Cardiovascular assessment and management. implications for preoperative clinical evaluation. *Minerva Anesthesiology*. 2015;2:225-233.

22. Port SC. 2014 esc/esa guidelines on noncardiac surgery: Cardiovascular assessment and management. *Journal of Nuclear Cardiology*. 2016;1-3. doi:10.1007/s12350-016-0641-x.

23. Kristensen SD. 2014 esc/esa guidelines on noncardiac surgery: Cardiovascular assessment and management. *Journal of Nuclear Cardiology*. 2016;1-3. doi:10.1007/s12350-016-0642-9.

24. Aert RCM van, Wicherts JM, Assen MALM van. Conducting meta-analyses based on p values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science*. 2016;11(5):713-729. doi:10.1177/1745691616650874.