International Journal of Epidemiology, 2015, 862–869
doi: 10.1093/ije/dyv114
Advance Access Publication Date: 13 July 2015
Original article

Spot the Discrepancy

Difficulty in detecting discrepancies in a clinical trial report: 260-reader evaluation

Graham D Cole,[1]* Matthew J Shun-Shin,[1] Alexandra N Nowbar,[2] Kevin G Buell,[2] Faisal Al-Mayahi,[2] David Zargaran,[2] Saliha Mahmood,[1] Bharpoor Singh,[2] Michael Mielewczik[1] and Darrel P Francis[1]

[1]International Centre for Circulatory Health and [2]School of Medicine, Imperial College London, London, UK

*Corresponding author. International Centre for Circulatory Health, National Heart and Lung Institute, Imperial College London, 59-61 North Wharf Road, London W2 1LA, UK. E-mail: g.cole@imperial.ac.uk

Accepted 9 April 2015

Abstract

Background: Scientific literature can contain errors. Discrepancies, defined as two or more statements or results that cannot both be true, may be a signal of problems with a trial report. In this study, we report how many discrepancies are detected by a large panel of readers examining a trial report containing a large number of discrepancies.

Methods: We approached a convenience sample of 343 journal readers in seven countries, and invited them in person to participate in a study. They were asked to examine the tables and figures of one published article for discrepancies. 260 participants agreed, ranging from medical students to professors. The discrepancies they identified were tabulated and counted. There were 39 different discrepancies identified. We evaluated the probability of discrepancy identification, and whether more time spent or greater participant experience as academic authors improved the ability to detect discrepancies.

Results: Overall, 95.3% of discrepancies were missed. Most participants (62%) were unable to find any discrepancies. Only 11.5% noticed more than 10% of the discrepancies. More discrepancies were noted by participants who spent more time on the task (Spearman's $\rho = 0.22$, $P < 0.01$), and those with more experience of publishing papers (Spearman's $\rho = 0.13$ with number of publications, $P = 0.04$).

Conclusions: Noticing discrepancies is difficult. Most readers miss most discrepancies even when asked specifically to look for them. The probability of a discrepancy evading an individual sensitized reader is 95%, making it important that, when problems are identified after publication, readers are able to communicate with each other. When made aware of discrepancies, the majority of readers support editorial action to correct the scientific record.

Key words: Peer review, retraction of publication, clinical governance, patient safety

Key Messages

- Even when directed to look for them, discrepancies in a trial report are difficult for readers to spot.
- Journal editors made aware of discrepancies in a published trial should not assume that other readers will be able to identify the discrepancies for themselves.
- When made aware of discrepancies, the majority of readers support editorial action to correct the scientific record.

Introduction

It is known that scientific literature contains errors.[1,2] We have noticed that clinical trial reports may sometimes contain discrepancies, i.e. two statements or results that cannot both be true. The impact of discrepancies on overall reliability is unproven, but there are examples where they were an accessible and early warning that the findings of the reports were unreliable[3] and that research failure had occurred.[4]

Previous studies have demonstrated that it is difficult for peer reviewers to spot problems when they are inserted as part of an experiment.[5–8]

This study addresses the difficulty readers face when looking for problems in a published article containing numerous naturally occurring discrepancies.

We asked participants to study 'The acute and long-term effects of intracoronary Stem cell Transplantation in 191 patients with chronic heARt failure: the STAR-heart study',[9] for two reasons. First, it was an article with numerous discrepancies of many different types, offering participants maximum opportunity to pick up problems. Second, it had an unusual pedigree, having completed peer review, publication, criticism, editorial re-evaluation, statistical re-review and subsequent exoneration. At the end of this process, the journal's decision was that, whereas it would relay news of a duplicate publication,[10] it was unable to share with readers the multitude of internal discrepancies or the contradiction with the alternative publication where the data presented in this paper as an observational study of 391 patients are presented identically but described as a blinded randomised controlled trial of 578 patients.[9,10]

This might be a suitable editorial approach if almost all readers can identify problems with the article. However, if readers cannot spot discrepancies, it may be more important for journals to bring those that are indeed spotted to the attention of readers.

In this paper we tested whether readers would be able to spot discrepancies. We also surveyed what readers would wish to happen when a paper is discovered to contain many discrepancies.

Methods

Sampling of readers

A total of 343 individuals across several countries were invited in person at academic conferences and places of academic work by study staff to review the article[9] for the purpose of this research study that we were conducting. The individuals approached were clinical or research doctors, medical students, undergraduate students at a scientific institution or scientific staff working in either a hospital environment or in industry. This was a convenience sample and we did not attempt to stratify for particular roles or experience. Participants provided verbal consent. We made it clear that their voluntary participation was in a research study. They were only told the identity of the paper to be reviewed once they were ready to examine it. Their responses were anonymous and no patients were involved, and therefore verbal consent was considered proportionate and reasonable. Participants' willingness to fill in the survey pro forma voluntarily was taken to indicate consent. Guidance indicated that research ethics committee approval is not required for such a study.

Detection of discrepancies

Participants were asked to read the paper and directed to 'examine the tables and figures of the results for discrepancies'. Beyond this, they were not directed to any feature. Each participant was asked to provide their age, sex, job role and numbers of publications they had authored or co-authored, and to note the approximate time they took for this task. No particular duration of study was suggested. Whereas they were guaranteed anonymity, they were also invited to volunteer their name and e-mail address on a detachable portion to assist any later audit. The data collection sheet is shown as Online Supplement 1, available as Supplementary data at *IJE* online.

Classification of discrepancies

From previous work,[3,11] we had identified four types of discrepancy present in the paper.

Impossible percentages

When describing a percentage of 200 patients, each patient represents 0.5%. Percentages such as 18.1% in Table 1 of the paper[9] (the number of controls with an RCX lesion) are not possible.

Arithmetical errors

For example when the VO_2 in the control group changes from a baseline of 1546 ml/min to 1539, the change is –7 and not –29.3 as written in Table 2 of the paper.[9]

Missed P-values

In the tables of the paper,[9] the convention appears to be that the changes that are statistically significant are asterisked as such and those that are not have no asterisk. Under this convention, it would be an error to mark a non-significant change as significant or to leave a significant change without an asterisk. In Table 2 of the paper, the O_2-pulse changes by $+0.52 \pm 2.1$ in the 191-patient treatment group compared with -0.9 ± 1.2 in the 200-patient control group. There is no asterisk and therefore this would be assumed to by readers to be not significant. However when the significance is calculated, it is $P < 0.00000000000000000001$.

Other discrepancies

These were other factual impossibilities, such as patients who had already died or were lost to follow-up on the survival analysis being documented as returning for follow-up, clinical assessment and investigations.

Participant feedback on their role in detecting discrepancies

After handing in their responses, participants were provided with annotated versions of the tables and figures displaying the discrepancies that prolonged analysis by the authors of the study had revealed.

Now aware of the extent and variety of discrepancies and problems, participants were invited to answer a series of questions (Online Supplement 2, available as Supplementary data at *IJE* online) about what they had been looking for and what they thought of the paper persisting without further action.

Data collection and analysis

Where an participant indicated multiple examples of the same type of discrepancy in a table, we scored them as having noticed all discrepancies of that type in that table.

The continuous data were not normally distributed, so relationships were tested with Spearman's rank correlation coefficient, and comparisons were conducted with the Mann-Whitney U-test.

Exclusion of 'missed *P*-values' from analysis

During peer review of this manuscript, it emerged that 'missed ***P***-values' (failure to asterisk significant changes when others significant changes were asterisked) did not fulfil our strict definition of a discrepancy unless the paper stated that a comparison was being made. Peer reviewers did not disagree that these 'missed ***P***-values' were serious problems with the paper. However, we concurred that, in hindsight, it was unfair to expect participants to identify these when briefed to look for discrepancies based on our definition. We were concerned that some participants might have noticed the 'missed ***P***-values' but judged them not to fulfil our definition. So as not to underestimate participant performance, we therefore removed this data from our analysis unless the paper specifically described the groups as comparable. However, the full list of problems in the paper is presented in Online Supplement 3, available as Supplementary data at *IJE* online.

Similarly, we agreed with the reviewers in hindsight that it was unfair to ask participants to spot that the output in Figure 3 of the paper was not generated from the SPSS survival software package as described, since this required some specialist knowledge. For example, standard Kaplan-Meier curves in SPSS start at time 0 (not time 1 year as plotted) and show steps at each event and not a smooth curve as plotted.

Results

Participant characteristics

In all, 343 individuals were approached and invited to take part and 260 individuals working in seven countries agreed to read the paper, a response rate of 76%. The characteristics of participants are shown in Table 1.

Participant performance

The publication contains 37 discrepancies that we were aware of before conducting the study, and additional problems not fulfilling our strict definition of a discrepancy. During this study, the 260 participants between them identified a further 2 discrepancies not noticed by the authorship team, giving 39 in total. There were therefore 10 140 (260×39) individual opportunities for discrepancy detection. In total, 474 (4.7%) of the potential discrepancies were identified; 161 participants (62%) did not find any of the 39 discrepancies. The number of discrepancies noted

Table 1. Characteristics of participants studying the paper. Data are provided either as number and percentages, or as median with interquartile range. Asterisk indicates that this was for the 96 (37%) participants with publications

Characteristic		Respondents
Role	Consultant/Professor	23 (9%)
	Post-Doctoral Scientist	7 (3%)
	Senior Medical Trainee	49 (19%)
	Junior Medical Trainee	24 (9%)
	Research Students	9 (3%)
	Medical Students	130 (50%)
	Other	12 (5%)
	Not Provided	6 (2%)
Age		23(21 to 30)
	Not Provided	16 (6%)
Gender	Male	162 (62%)
	Female	93 (36%)
	Not Provided	5 (2%)
Number of Publications*		4.5 (2 to 17)
	Not Provided	16 (6%)
Time Spent Reading Paper (mins)		20 (15 to 30)
	Not Provided	94 (36%)

by individuals ranged from 0 to 26 with median 0 [interquartile range (IQR) 0 to 1]. Only 30 (11.5%) of participants found more than 10% of the discrepancies.

Predictors of participant performance

Spending more time was associated with identifying more discrepancies (Spearman's $\rho = 0.22$, $P < 0.01$). There was a weak correlation between the number of publications and the number of discrepancies identified ($\rho = 0.13$, $P = 0.04$).

Half of our participants were medical students. Participants who were not medical students picked up more discrepancies than medical students (median 0 discrepancies, IQR 0 to 1, vs median 0, IQR 0 to 1, $P < 0.01$).

Some discrepancies were detected by more participants than others

Some of the discrepancies were recognized much more frequently than others, as shown in Figure 1. The most frequently noted discrepancy was that the baseline ejection fraction was significantly different between the treatment and control groups, but this still evaded 82% of participants. None of our 260 participants detected that the paper implied the existence of a patient with a negative (and therefore impossible) New York Heart Association functional class of −24.8.[3]

Self-reported focus of attention

Despite participants being specifically directed to look at the figures and tables for discrepancies, when later asked what they had actually been doing, many reported looking at the results for possible conclusions or comparing the two groups rather than looking specifically for discrepancies. Only 78 of the 260 participants (30%) described undertaking discrepancy-seeking behaviour for both Table 1 and Table 2 of the paper.[9]

Participant opinion of the paper once shown the discrepancies

Having been shown the extent of the discrepancies, participants were asked whether the conclusions of the study were correct. Of the 65% responding, the most common response was that the paper's conclusion was not correct (49%); 29% gave answers that attempted to weigh up different aspects of the paper without giving a clear overall opinion. Only 22% of participants gave an answer indicating that they thought the conclusion was correct.

We told participants that the paper persisted in the literature despite the journal being made aware of the discrepancies, with no plan for retraction or a note of concern. We offered participants a free text response to list what they thought should now happen, and 154 (59%) gave an opinion. Only 10 (6.5% of those responding) suggested that the current situation be allowed to persist without further action. These participants were outnumbered over 5:1 by the 51 (33%) who suggested the paper be retracted. Other common suggestions were that the authors be asked further questions (19%), a third-party investigation be conducted (14%), reform to the peer review system (8%), re-re-review of the paper (8%) or the authors or journal be penalised (10%).

Discussion

Our study shows that it is difficult for readers to detect discrepancies in a paper, even when these are numerous. Previous work has shown that peer reviewers frequently do not pick up on weaknesses in a trial's description.[12] They also have difficulty detecting errors experimentally injected into articles.[6] Tables and figures are a particular challenge because it seems that peer reviewers rarely notice problems in them.[8] In our study, we specifically directed the participants to look for discrepancies and to focus on the tables and figures. Therefore, without such prompting, real-life rates of detection of discrepancies may be lower.

Discrepancies do not automatically mean that data have been deliberately misrepresented. They can occur for many

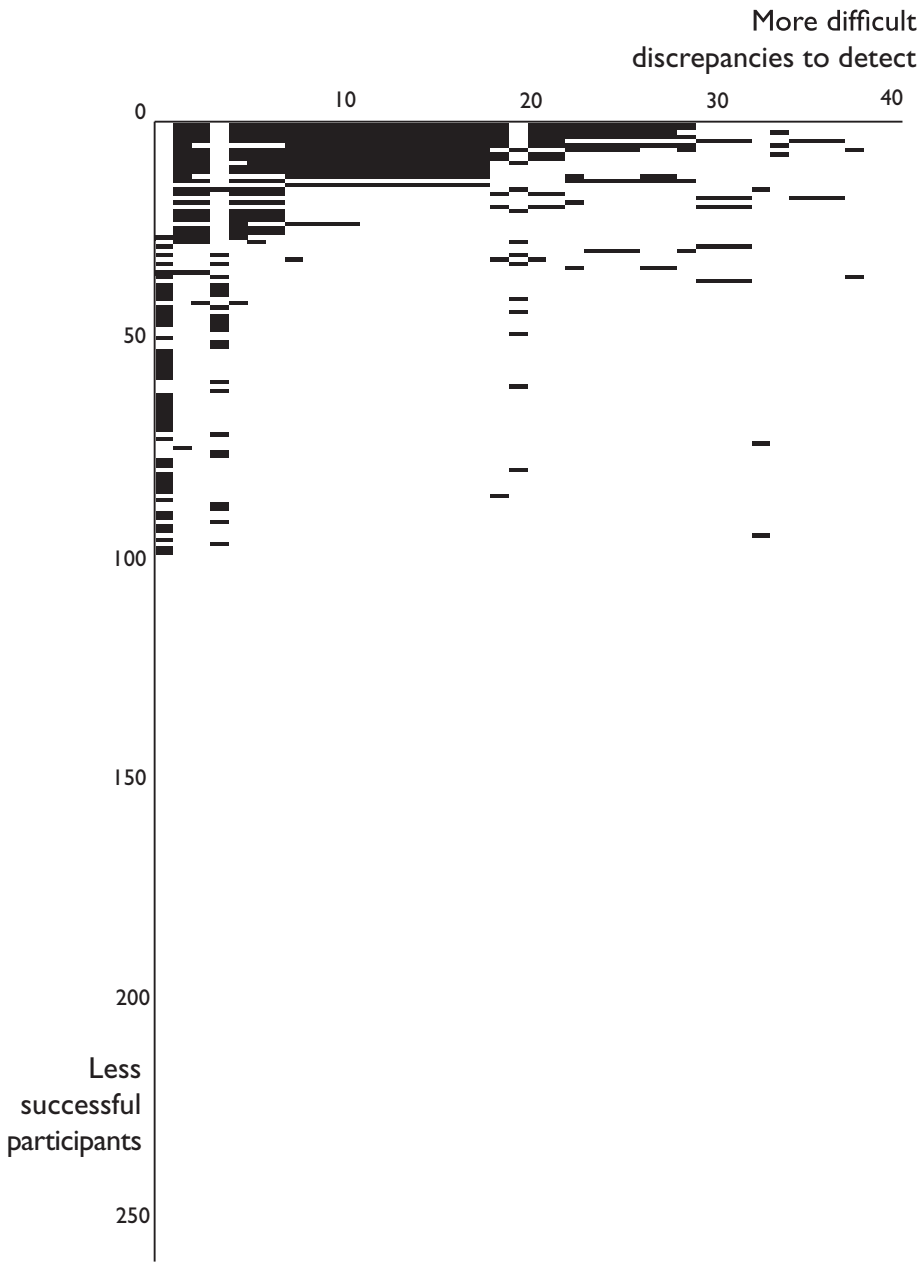

Figure 1. Spectrum of discrepancy recognition. For any research paper with discrepancies, this plot tests the hypothesis that each reader is capable of finding them on their own, and therefore does not need the discrepancies to be communicated via the journal. Each black area represents a detected discrepancy. Each of the 39 columns represents a different discrepancy and has been arranged by decreasing chance of detection by participants. Each of the 260 rows represents a participant and has been arranged from most successful in detecting discrepancies at the top to least successful at the bottom. If all readers were capable of detecting all discrepancies independently, the entire diagram would be black. The concentration of black areas in the top-left corner indicates that some discrepancies were much easier to find than others.

reasons.[13] For example, one innocent explanation for a mismatched percentage is that the denominator of a proportion may be different from the entire group size when data are missing in some patients, and the authors have forgotten to state this. Alternatively, there may have been an error in the numerator, or in the calculation of the percentage. A more disappointing possibility is that patients have been moved between groups. From the trial report alone it is not possible to know which is the case.

Readers might think that some discrepancies are more serious than others. As authors of this study, we think this too, but cannot agree on the hierarchy of seriousness among ourselves or with others. For example, the 391-patient STAR-heart observational study[9] has data numerically identical in every way to the 578-patient randomized controlled trial BEST-heart trial.[10] However, when evidence of this was shown to the journal editors, they did not consider this serious and were satisfied to issue a note of partial duplicate publication.[14]

A far more prevalent problem than inaccurate reporting is methodological weakness,[15,16] for example over-attribution of differences in outcomes between observational groups to differences in therapeutic choices. However, whether inferences have been made too strongly from weak studies is a qualitative judgement that can be open to debate. In our present study we addressed only plain discrepancies: pairs of statements that cannot both be true.

This study indicates that, even when explicitly asked to identify discrepancies, an individual discrepancy has a 95% chance of escaping the notice of a reader. Even when readers notice a single discrepancy, they frequently miss many others present in the same paper. There were weak tendencies for those who spent more time and those who had more experience of publishing research to detect more discrepancies. However, even those who had published research, and who spent the median time or greater assessing the paper, still only picked up 11.5% of the discrepancies. Even though participants who were not medical students picked up more discrepancies than medical students, the median number of discrepancies found by participants who were not medical students was still 0.

When made aware of the scale of discrepancies, in contrast to the belief of some journals, many more readers support notification of other readers and retraction than wish the paper to persist in the literature without further action.

Should journals expect individual readers to detect all discrepancies themselves?

Scientists may assume that a quality control process takes place before publication. This may be the case, but our correspondence[3] indicates that when this fails, individual readers themselves may have to rely on their own ability to recognize discrepancies. Not all journals are able to provide a forum for readers to communicate their concerns to each other. Therefore, even though the community of readers can be very large, they are unable to help each other by building up a complete picture of the factual impossibilities.

In this study we directed participants to look specifically for discrepancies. When readers normally read a paper, there is no such direction, and therefore the chance of recognizing them is likely to be even more grim. Journal editors hope and assume that readers will check for discrepancies, whereas readers, unaware of this responsibility, hope and assume that journal editors have already done so.

Institutions commonly espouse careful use of public and charitable resources for research. They employ staff who rely on journals to communicate globally. However, when their staff find discrepancies and report them to journals, they are not communicated to others. Is it good value to pay journals for access to information that journals know is incorrect?

Post-publication processes

Journals differ in the opportunities offered for post-publication dialectic. Our experience in this therapeutic field is that the *British Medical Journal*'s rapid response system[11,17] allowed experts to draw attention to other studies with numerous discrepancies. In contrast, the *Journal of the American College of Cardiology* has an unbreakable limit of one round of questioning[18] per paper. This policy preserves as a mystery what happened to the radionuclide primary endpoint data[19,20] or how a group mean can increase by +7.0 when the mean increment per patient is displayed as +5.4.[18]

The arrival of platforms for post-publication discussion provides better opportunities, because journals cannot block them. For example, the American National Institutes of Health (NIH) provides the Pubmed Commons platform, and PubPeer is making a mark[21] as a non-governmental alternative.

Critical appraisal of an article involves far more important aspects than detecting discrepancies. For example, readers should be aware of the limitations of making therapeutic decisions based on observational comparisons[22,23] and be aware of the need for appropriate statistical testing. However, when readers note discrepancies in a trial, it would be helpful if they made them available to other readers because most readers will not notice most discrepancies in the normal course of events. Such an approach might better leverage more extensive education in critical appraisal. It is notable that many of the problems in this paper arose in tables, which another study has identified as a difficult part of the peer review process.[8]

It is likely that methodological experts would pick up many of these discrepancies while scrutinizing the trial for the commoner failings of design. However, it may not be practical to engage such experts to check every published study. Instead, it may be more cost-efficient for journals to facilitate readers to relay such notifications to each other.

Clinical implications

Even when readers take the time to write letters asking questions of authors, many go unanswered,[24] mirroring our experience.[3,25] Clinical guidelines include trials subject to unanswered correspondence.[24] When concerns arise regarding clinical research, it is imperative that the

message can be shared with others to promote careful scrutiny and avoid harm to patients. For years, cardiologists across Europe may have unwittingly done harm to patients undergoing non-cardiac surgery. Their mistake was nothing more than following European Society of Cardiology (ESC) guidelines advocating perioperative beta blockade.[26] The DECREASE family of clinical trials that formed the bedrock of these recommendations are now suspected to be either fabricated or fictitious.[27–31] We have shown that the remaining credible trials show perioperative beta-blockade to be associated with harm.[32] Through sanctioned guidelines, this research may have cost, according to ESC expert formulae,[31] thousands of lives.[33]

The conclusions of the trial we study in this paper are now incorporated into a meta-analysis entitled 'Adult bone marrow cell therapy improves survival and induces long-term improvement in cardiac parameters'. This meta-analysis[34] undertook a quality assessment which, like 62% of readers, showed no sign of seeing any discrepancies. In other fields, detailed examination of a trial report can identify serious problems not revealed by systematic reviews using checklists to assess quality.[35] Unless journals can facilitate post-publication discussion by readers, trials with serious discrepancies can percolate via meta-analysis through to clinical practice guidelines and ultimately put patients in danger.

Limitations

Our participants were a convenience sample rather than a systematically targeted group. We do not know whether different groups would fare differently. Many of our participants were junior and inexperienced. Experience of publishing research had a modest positive association with noticing discrepancies. However, junior and inexperienced people do read papers and, since this journal is unable to relay to them the problems in this paper, this is the magnitude of the challenge they face.

We only studied a single paper and therefore a single area. We did this because it contains a large number of discrepancies and therefore discrepancies should be easy to find in it. If we had asked participants to address papers with fewer discrepancies, then the number picked up might have been even lower.

Conclusions

We found that 95% of discrepancies go unnoticed even by readers specifically asked to look for them and directed to the figures and tables. Currently, discrepancies reported to journals[3] are not always relayed to readers as they come to light. Journals should tell readers that each must do their own discrepancy detection personally. Even so, individuals will miss most discrepancies. Because individuals find only a small fraction of discrepancies, it is crucial that a forum exists for readers to pool their observations. Our experience is that not all journals are ready to provide this.

Readers becoming aware of many discrepancies in an article disapprove of it persisting in the journal with no warning given to other readers. Guideline writing committees may not notice discrepancies reported and published,[24] but certainly cannot notice discrepancies reported by readers to editorial boards and then buried.

The number of readers required to identify a particular discrepancy may be hundreds or thousands. Even minor discrepancies should not be neglected, as they may be the tip of an error iceberg.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

G.D.C. and M.J.S. are British Heart Foundation Clinical Research Training Fellows (FS/12/12/2924 and FS/14/27/30752). D.P.F. is a British Heart Foundation Senior Fellow (FS/10/038). The funder had no role in devizing, conducting, analysing or reporting this study.

Acknowledgements

We are grateful to the 260 readers who read this paper with many discrepancies.The authors are grateful for infrastructural support from the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London.
The senior author (D.P.F.) is guarantor.

Conflict of interest: We have no conflicts of interest to declare.

References

1. Altman DG. Statistics in medical journals. *Stat Med* 1982;**1**: 59–71.
2. Altman DG. The scandal of poor medical research. *BMJ* 1994; **308**:283–284.
3. Francis DP, Mielewczik M, Zargaran D, Cole GD. Autologous bone marrow stem cell therapy in heart disease: discrepancies and contradictions. *Int J Cardiol* 2013;**168**: 3381-403.
4. Abbot A. Evidence of misconduct found against cardiologist. *Nature News Blog*. 24 February 2014. http://blogs.nature.com/news/2014/02/evidence-of-misconduct-cardiologist.html (17 March 2014, date last accessed).
5. Nylenna M, Riis P, Karlsson Y. Multiple blinded reviews of the same two manuscripts. Effects of referee characteristics and publication language. *JAMA* 1994;**272**:149–151.
6. Baxt WG, Waeckerle JF, Berlin JA, Callaham ML. Who reviews the reviewers? Feasibility of using a fictitious manuscript to

evaluate peer reviewer performance. *Ann Emerg Med* 1998;**32**: 310–17.
7. Godlee F, Gale CR, Martyn CN. Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomised controlled trial. *JAMA* 1998;**280**:237–40.
8. Schroter S, Black N, Evans S, Godlee F, Osorio L, Smith R. What errors do peer reviewers detect, and does training improve their ability to detect them? *J R Soc Med* 2008;**101**:507–514.
9. Strauer B-E, Yousef M, Schannwell CM. The acute and long-term effects of intracoronary Stem cell Transplantation in 191 patients with chronic heARt failure: the STAR-heart study. *Eur J Heart Fail* 2010;**12**:721–29.
10. Strauer BE, Yousef M, Schannwell MC. The BEST-Heart-Study: the acute and long term benefit of autologous stem cell transplantation in 289 patients with chronic heart failure. In: Strauer BE, Ott G, Schannwell CM (eds). Adulte Stammzellen [in German]. Dusseldorf, Germany: Dusseldorf University Press, 2009.
11. Nowbar AN, Mielewczik M, Karavassilis M *et al*. Discrepancies in autologous bone marrow stem cell trials and enhancement of ejection fraction (DAMASCENE): weighted regression and meta-analysis. *BMJ* 2014;**348**:g2688.
12. Hopewell S, Collins GS, Boutron I *et al*. Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. *BMJ* 2014;**349**: g4145.
13. Moyé L. DAMASCENE and meta-ecological research: a bridge too far. *Circ Res* 2014;**115**:484–87.
14. Strauer B-E, Yousef M, Schannwell CM. Corrigendum to 'The acute and long-term effects of intracoronary Stem cell Transplantation in 191patients with chronic heARt failure: the STAR-heart study [*Eur J Heart Fail* 2010;**12:**721–29]. *Eur J Heart Fail* 2013;**15**:360.
15. Howard JP, Cole GD, Sievert H *et al*. Unintentional overestimation of an expected antihypertensive effect in drug and device trials: mechanisms and solutions. *Int J Cardiol* 2014;**172**: 29–35.
16. Rosen MR, Myerburg RJ, Francis DP *et al*. Translating stem cell research to cardiac disease therapies: pitfalls and prospects for improvement. *J Am Coll Cardiol* 2014;**64**:922–37.
17. Moye L. Rapid response to DAMASCENE. 2014. http://www.bmj.com/content/348/bmj.g2688/rr/699167 (16 December 2014, date last accessed).
18. Mielewczik M, Cole GD, Nowbar AN *et al*. The C-CURE Randomized Clinical Trial (Cardiopoietic stem Cell therapy in heart failURE). *J Am Coll Cardiol* 2013;**62**:2453.
19. World Health Organization. International Clinical Trials Registry Platform ID EUCTR2007-007699-40-BE. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2007-007699-40-BE (17 December 2014, date last accessed).
20. European Union. Clinical Trials Register. ID 2007-007699-40. https://www.clinicaltrialsregister.eu/ctr-search/trial/2007-007699-40/BE (17 December 2014, date last accessed).
21. Oransky, I. Author with six recent corrections retracts JBC paper questioned on PubPeer. http://retractionwatch.com/2013/08/26/author-with-six-recent-corrections-retracts-jbc-paper-questioned-on-pubpeer/ (16 December 2014, date last accessed).
22. Sen S, Davies JE, Malik IS *et al*. Why does primary angioplasty not work in registries? Quantifying the susceptibility of real-world comparative effectiveness data to allocation bias. *Circ Cardiovasc Qual Outcomes* 2012;**5**:759–66.
23. Nijjer SS, Pabari PA, Stegemann B *et al*. The limit of plausibility for predictors of response: application to biventricular pacing. *JACC Cardiovasc Imaging* 2012;**5**:1046–65.
24. Horton R. Postpublication criticism and the shaping of clinical knowledge. *JAMA* 2002;**287**:2843–47.
25. Cole G, Francis DP. Comparable or STAR-heartlingly different left ventricular ejection fraction at baseline? *Eur J Heart Fail* 2011;**13**:234.
26. Task Force for Preoperative Cardiac Risk Assessment and Perioperative Cardiac Management in Non-cardiac Surgery, European Society of Cardiology (ESC), Poldermans D *et al*. Guidelines for pre-operative cardiac risk assessment and perioperative cardiac management in non-cardiac surgery. *Eur Heart J* 2009;**30**:2769–812.
27. Erasmus Medical Centre. 2011. http://www.erasmusmc.nl/research/over-research/wetenschappelijke-integriteit/rapport/?lang=en (21 October 2013, date last accessed).
28. Erasmus Medical College Press Release. 2011. http://www.erasmusmc.nl/corp_home/corp_news-center/2011/2011-11/ontslag.hoogleraar/?lang=en (21 October 2013, date last accessed).
29. Erasmus Medical Centre. 2012. Report on the 2012 follow-up investigation of possible breaches of academic integrity. http://www.erasmusmc.nl/5663/135857/3675250/3706798/Integrity_report_2012-10.pdf?lang=en&lang=en (21 October 2013, date last accessed).
30. Chopra V, Eagle KA. Perioperative mischief: the price of academic misconduct. *Am J Med* 2012;**125**:953–55.
31. Cole GD, Francis DP. Perioperative β blockade: guidelines do not reflect the problems with the evidence from the DECREASE trials. *BMJ* 2014;**349**:g5210.
32. Bouri S, Shun-Shin MJ, Cole GD *et al*. Meta-analysis of secure randomised controlled trials of β-blockade to prevent perioperative death in non-cardiac surgery. *Heart* 2014;**100**:456–64.
33. Husten, L. Medicine Or Mass Murder? Guideline Based on Discredited Research May Have Caused 800,000 Deaths In Europe Over The Last 5 Years. http://www.forbes.com/sites/larryhusten/2014/01/15/medicine-or-mass-murder-guideline-based-on-discredited-research-may-have-caused-800000-deaths-in-europe-over-the-last-5-years/ (17 March 2014, date last accessed).
34. Jeevanantham V, Butler M, Saad A, Abdel-Latif A, Zuba-Surma EK, Dawn B. Adult bone marrow cell therapy improves survival and induces long-term improvement in cardiac parameters: a systematic review and meta-analysis. *Circulation* 2012;**126**: 551–68.
35. Hirji KF. No short-cut in assessing trial quality: a case study. *Trials* 2009;**10**:1.