Discrepancies in clinical trial reports: frequency in retracted reports versus unretracted reports

[1] Graham D Cole MA MB BChir MRCP

[1] Alexandra N Nowbar BSc

[1] Michael Mielewczik Dr. sc. ETH, Dipl.-Biol.

[1] Matthew J Shun-Shin BA BMBCh MRCP

[1] Darrel P Francis MA MB BChir MD FRCP

1 International Centre for Circulatory Health, National Heart and Lung Institute, Imperial College London, W2 1LA

Correspondence:
Dr Graham D Cole
International Centre for Circulatory Health
National Heart and Lung Institute
Imperial College London
59-61 North Wharf Road
London W2 1LA, UK

Tel: +44 207 594 1093
Fax: +44 208 082 5109
Email: g.cole@imperial.ac.uk

Word count: 3879

Keywords : Retraction of Publication, Scientific misconduct, Clinical governance, Patient safety, Research Integrity

Abstract

Background

Readers of clinical trials have difficulty judging whether they are reliable. Publications sometimes contain discrepancies: mathematically or logically contradictory statements. These include percentages that are impossible for the stated number of patients, incorrect but theoretically possible percentages, impossible summary statistics, other arithmetical errors, and p-values incorrectly implied non-significant. However, whether such discrepancies are meaningful has been controversial.

Method

50 retracted clinical trial reports were randomly selected from PubMed. The preceding clinical trial report in the same journal acted as a control. All traces of retraction were removed. Three scientists, blinded to the retraction status of individual reports, reviewed all 100 trial reports for discrepancies. Discrepancies were pooled and cross-checked before being counted into pre-specified categories. Only then was the retraction status unblinded for analysis.

Results

We found 479 discrepancies in the 100 trial reports, of which 348 were in the 50 retracted reports and 131 in the 50 unretracted reports. The total number of discrepancies in the 50 retracted reports was 2.7-fold higher than in the unretracted reports. On average, individual retracted reports had a greater number of discrepancies (median 4, IQR 2 to 8.75) than unretracted (median 0, IQR 0 to 5), $p<0.01$. Papers with a discrepancy were significantly more likely to be retracted than those without a discrepancy (OR 5.7, 95% CI 2.2 to 14.5, $p<0.001$).

In particular, three types of discrepancy arose significantly more frequently in retracted than unretracted reports: factual discrepancies ($p<0.01$), arithmetical errors ($p=0.01$) and missed p-values ($p=0.02$).

Conclusions

Discrepancies in published trial reports should no longer be assumed to be unimportant. We found that scientists, blinded to retraction status and with no specialist skill in the field, identify significantly more discrepancies in retracted than unretracted clinical trial reports. Discrepancies may be an early and accessible signal of unreliability in clinical trial reports.

Introduction

Landmark science cannot always be replicated independently.(1,2,3) Erroneous research is not uncommon(4,5) and wastes intellectual and financial resources. More importantly, the incorrect results may spawn further clinical research which needlessly draws more patients into trials that would not have been initiated had the original research been reported correctly. In some cases, insecure clinical trials can harm patients when doctors implement their findings in good faith.(6,7,8)

In the field of bone-marrow stem cell therapy for heart disease, for example, readers are faced with a wide spectrum of conflicting effect sizes which conventional meta-analyses have been unable to explain. In that field, we have recently reported that the numbers of mathematical or logical discrepancies per trial are the strongest determinant of the effect size reported by the trial.(9) However, currently, such discrepancies are assumed by some journals to be unimportant and not worth highlighting to readers.(10) Reaction to identification of hundreds of discrepancies in only one field varied from interest(11) to criticism that the entire analysis should be "set aside" and that discrepancies should be routinely accepted as insignificant "flubs".(12)

Although retractions are increasing,(13) the level remains far lower than the rate of erroneous research(5) implying that the literature may be burdened by a substantial proportion of findings that are insecure but unretracted and therefore unrecognized. If discrepancies are more common in retracted studies than unretracted studies, they might represent an accessible signal of concern for readers.

We therefore investigated whether discrepancies are more prevalent in retracted than adjacent unretracted reports in the same journals.

Methods

We undertook a blinded case-control study. We identified discrepancies in randomly-selected retracted clinical trials, using, in each case, the preceding unretracted trial in the same journal as the control. We used the same journal as this has been identified as a major source of variation in retraction rates.(14) Annotations of retraction were removed, the studies were presented in random order to three scientists, who were asked to remain blinded to retraction status.

A PubMed search was conducted in December 2012 for the "retracted publication" publication type and limited to Clinical Trials, with no restriction on publication date. A computer random number generator (Microsoft Excel RAND function) was used to select members of this set until 50 had been selected. For each of these trials, a paired control trial was also selected, defined as the unretracted clinical trial, in the same journal, whose PubMed accession sequence was immediately preceding the retracted trial. Watermarks of retraction were removed. The resulting 100 trials were given random sequence numbers between 1 and 100. We decided on a study size of 100 trial reports as a manageable number that could be studied by three scientists, given our previous experience examining reports for discrepancies.(9) The PDFs were presented to 3 scientists (GDC, ANN, MM) unaware of individual retraction status and asked to refrain from finding this out. Each independently identified factual or mathematical discrepancies without recourse to specialist knowledge.

Candidate discrepancies proposed by each scientist were pooled and duplicate candidates removed. All three scientists, joined by a fourth senior scientist (DPF), then examined all unique candidate discrepancies and gave an opinion on their individual validity as a discrepancy. At this

stage, conferring was allowed. Discrepancies were only accepted as valid if agreed as discrepancies by all four scientists.

Categories of discrepancy are shown, along with simple examples, in Table 1.

Table 1. Categories of discrepancy.

Name	Description	Example	Nature of Discrepancy
Possible Percentages	A percentage of a group of patients that does not match the subset count displayed but could fit a different subset count.	23 of 57 (42%) of patients in intervention group were taking an ACE inhibitor.(15)	23/57 is 40%. 42% of 57 patients is possible, but is 24 patients.
Impossible Percentages	A percentage of a group of patients that cannot exist without fractional patients.	31.2% of 200 patients had an infarct in the right coronary artery.(16)	Each patient represents 0.5% of the group. Percentages must be multiples of 0.5%. 31.2% is not.
Factual Discrepancies	Two statements that cannot both be true.	Abstract: Base excess was 1.04±0.3 in balanced group at baseline.(17) Results: Base excess was	Abstract and Results are mutually contradictory.

		1.16±0.3 in balanced group at baseline.(17)	
Impossible Summary Statistics	Summary statistics (mean, median, range, SD) that are not possible based on the data presented.	Median ICU stay in Unbalanced Group is 13 days. ICU stay ranged from 14 to 444 days.(17)	Median must lie within range.
Arithmetical Errors	Arithmetical errors such as subgroups that do not add up to the total parent group, or differences between before and after measurements that do not match the documented change.	Three subgroups of size 5, 5 and 6 patients received different doses of treatment. Total number treated 15.(18)	The three subgroups add to 16 patients, but the total is said to be 15.
Missed p-values	Two groups which are significantly different but are implied to be non-different (either explicitly or by omission of a symbol when other	Baseline ejection fraction in two groups of 29.4±12.7 (n=191) and 36.1±13.8 (n=200) described as comparable.(16)	These two groups are significantly different (p=0.0000006).

	comparisons are marked).		

Statistical Analysis

Descriptive statistics

The study was then unblinded and the reports re-paired. Overall discrepancy counts, and overall counts for the different categories, were compared between the 50 retracted and the 50 unretracted reports using the Wilcoxon matched-pairs signed-rank test. Odds ratios and their confidence intervals were calculated for comparisons between retracted and unretracted studies and the presence or absence of discrepancies.(19)

Regression analysis

The number of discrepancies between retracted and unretracted based could be driven by an extreme number of discrepancies in some retracted papers. Taking this into consideration, in a post-hoc analysis, we quantified the association between retraction status and the number of discrepancies by modelling the number of discrepancies using a zero-inflated, negative binomial model. We also used this model to consider the impact on discrepancy counts of retraction status, year of publication, citations of the trial report and journal impact factor. Regression coefficients are presented as incidence rate ratios (IRR) for the binomial component and odds ratios (OR) for the excess zero component.

Sensitivity and specificity analysis

We calculated sensitivity and specificity of detecting retraction for a range of cut-offs of discrepancy count, and the odds ratio for retraction above these cut-offs.

Statistical analysis was undertaken using "The R project for statistical computing"(20) (code shown in Online Supplement 1) with Figures prepared using "ggplot2".(21)

Data sharing

The complete list of identified discrepancies is shown in Online Supplement 3. These are freely available from us in editable form should any readers request it. We welcome and will make public any corrections or updates from readers.

We also provide the raw data in Online Supplement 6 to permit re-analysis of our data.

Results

Trial Reports

The search yielded 263 retracted reports of clinical trials published between 1983 and 2012, of which 50 were randomly selected. 24 (48%) were retracted for misconduct, 8 (16%) for errors, 7 (14%) for plagiarism, 5 (10%) for duplication, 1 (2%) for copyright issues. In 5 (10%), the reason was not stated or unclear. The PubMed IDs and number of discrepancies found in each report are provided in Online Supplement 2.

The trial reports and the discrepancies they contain are listed in Online Supplement 3. To allow readers to appreciate the findings of our study without necessarily seeing the identities of the trials or authors, each trial report is referenced by an R-code (R1 to R100). Nevertheless, and only to ensure verifiability, the discrepancies can be viewed in the original reports by entering the Pubmed IDs found in Online Resource 2 at www.pubmed.org.

Overall Discrepancy Counts

We found 479 discrepancies in the 100 trial reports, of which 348 were in the 50 retracted reports and 131 in the 50 unretracted reports. The overall number of discrepancies were 2.7-fold higher for the 50 retracted reports than the 50 unretracted reports. Individual report discrepancy counts were higher in retracted (median 4, IQR 2 to 8.75) than unretracted reports (median 0, IQR 0 to 5, $p<0.01$), as shown in Figure 1.

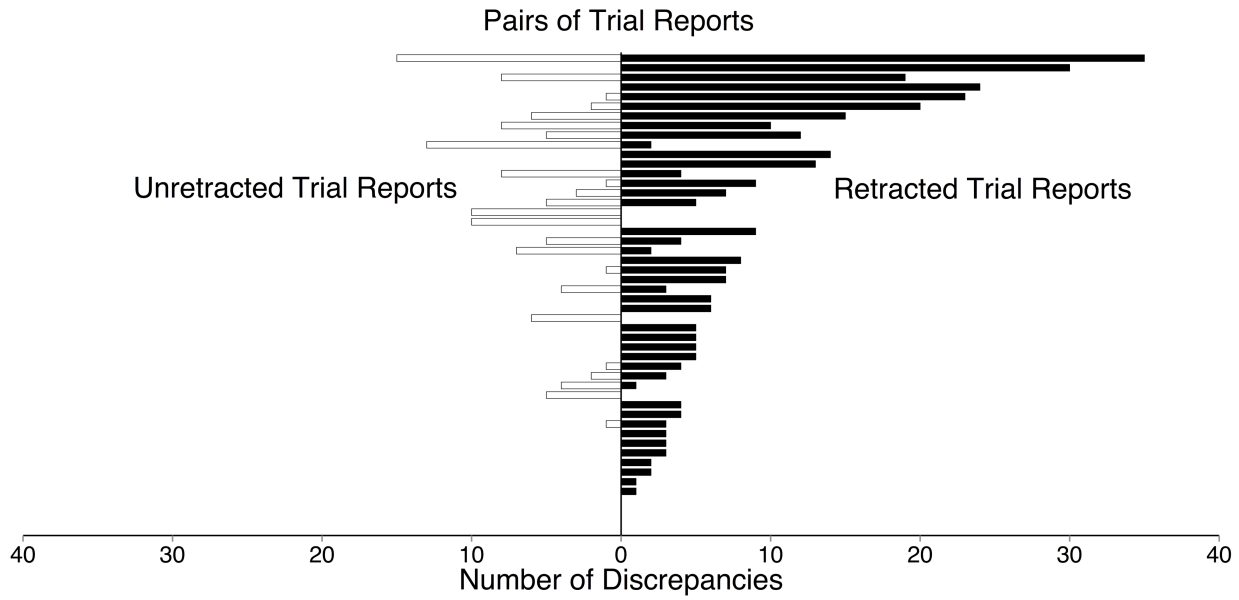


Figure 1. Discrepancy counts. Paired comparison between retracted reports (right) and unretracted control counterparts (left) for all 100 reports studied. Each bar represents the number of discrepancies in one trial. Pairs are ordered by total number of discrepancies in the pair, with those with most discrepancies at the bottom.

42/50 (84%) of retracted trials and 24/50 (48%) of unretracted trials contained discrepancies. In the 8 retracted trials with no discrepancies, the reason for retraction was misconduct in 4 cases, error in 2 cases and duplication in 2 cases.

Papers with a discrepancy were significantly more likely to be retracted than those without a discrepancy (OR 5.7, 95% CI 2.2 to 14.5, $p<0.001$).

Regression analysis

We considered the number of discrepancies in trial reports to be broadly a negative binomial distribution but with a certain excess proportion of reports with zero discrepancies. We therefore used a zero-inflated negative binomial regression to investigate the relationship between the number of discrepancies and retraction status, year, impact factor of the journal, and number of citations.

Retracted papers were more likely to have discrepancies, and more of them. In the formal analysis, the number of excess zeroes showed a significant relationship to retraction status. In this model retracted reports were less likely to have zero discrepancies (OR 0.14, 95% CI 0.03 to 0.67, p=0.01). No significant association was seen in relation to the year of publication, impact factor, and number of citations (table 2). This same pattern was seen in a univariable analysis (OR 0.11, 95% CI 0.02 to 0.79, p=0.03).

Similarly, the number of discrepancies was significantly related to retraction status. Retracted reports had significantly higher number of discrepancy counts as compared to unretracted reports (IRR 1.79, 95% CI 1.07 to 2.99, p=0.03). This same pattern was seen in a univariable analysis (IRR 1.62, 95% CI 1.07 to 2.69, p=0.06). No significant association was seen in relation to the year of publication, impact factor, and number of citations (table 2).

Table 2. Zero-inflated negative binomial analysis of discrepancies, retraction status, year of publication, journal impact factor and trial report citations. Only retraction status was significantly associated in both the excess zero components (negatively, in that trial reports are less likely to have zero discrepancies) and binomial (positively, in that retracted trial reports are more components of the model.

(a) Univariable analysis

Binomial Component	**Incidence Rate Ratio**	**95% Confidence Interval**	**p-value**
Intercept	4.65	(3.28 to 7.15)	<0.01
Retraction	1.62	(1.07 to 2.69)	0.06
Excess Zero Component	**OR**	**95% Confidence Interval**	**p-value**
Intercept	0.78	(0.42 to 1.62)	0.50
Retraction	0.11	(0.02 to 0.79)	0.03

(b) Multivariable analysis

Binomial Component	**Incidence Rate Ratio**	**95% Confidence Interval**	**p-value**
Intercept	2.45	(0.90 to 6.65)	0.08
Retraction	1.79	(1.07 to 2.99)	0.03
Year	1.05	(0.99 to 1.11)	0.11
Impact Factor	0.95	(0.90 to 1.00)	0.07
Citations	1.00	(1.00 to 1.01)	0.24
Excess Zero Component	**OR**	**95% Confidence Interval**	**p-value**
Intercept	4.22	(0.41 to 43.2)	0.23
Retraction	0.14	(0.03 to 0.67)	0.01
Year	0.88	(0.76 to 1.02)	0.08
Impact Factor	0.98	(0.83 to 1.16)	0.79
Citations	1.01	(1.00 to 1.01)	0.26

Types of Discrepancy

Some pre-specified discrepancy types were themselves significantly more likely in retracted trials, as shown in Figure 2. These were factual discrepancies (median 1, IQR 0 to 3 versus

median 0, IQR 0 to 0, p<0.01), arithmetical errors (median 0, IQR 0 to 0, range 0 to 2 versus median 0, IQR 0 to 0, range 0 to 6, p=0.01) and missed p-values (median 0, IQR 0 to 0, range 0 to 12 versus median 0, IQR 0 to 0, range 0 to 0, p=0.02). For types of discrepancy that did not show a statistically significant difference, the direction of the trend was in each case towards more discrepancies in the retracted papers.

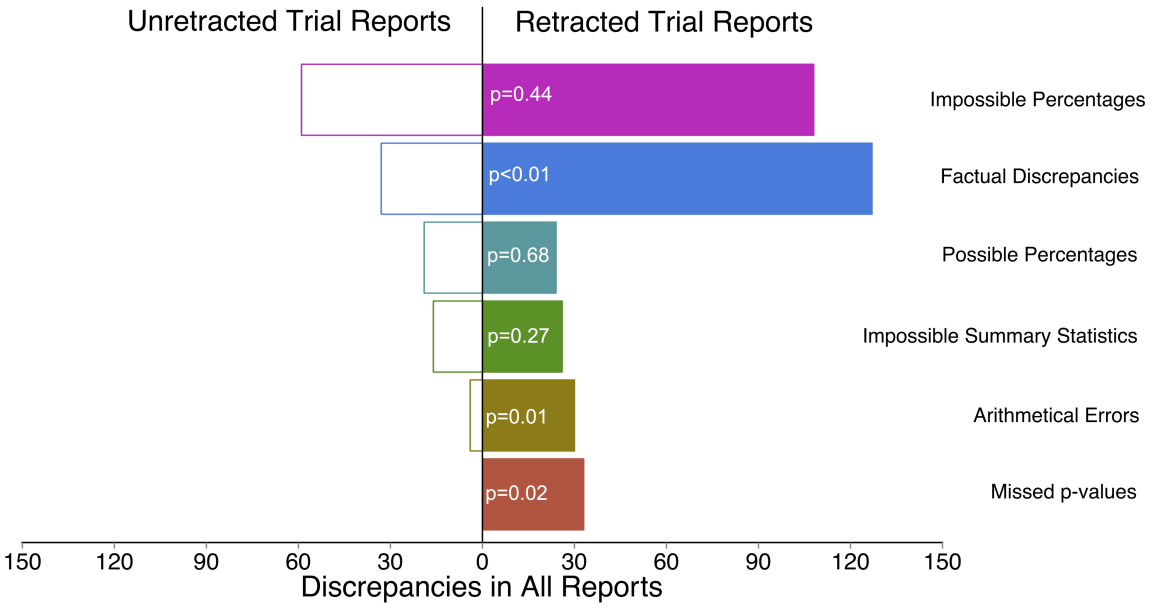


Figure 2. Total number of discrepancies by type across all 100 trial reports. Each bar represents the total number of one type of discrepancy in 50 trial reports. White bars are those found in unretracted control reports. Coloured bars are those found in retracted reports.

Sensitivity and Specificity

We considered whether a particular discrepancy threshold identified retracted reports by performing a sensitivity-specificity analysis on all 100 trial reports, as shown in Figure 3.

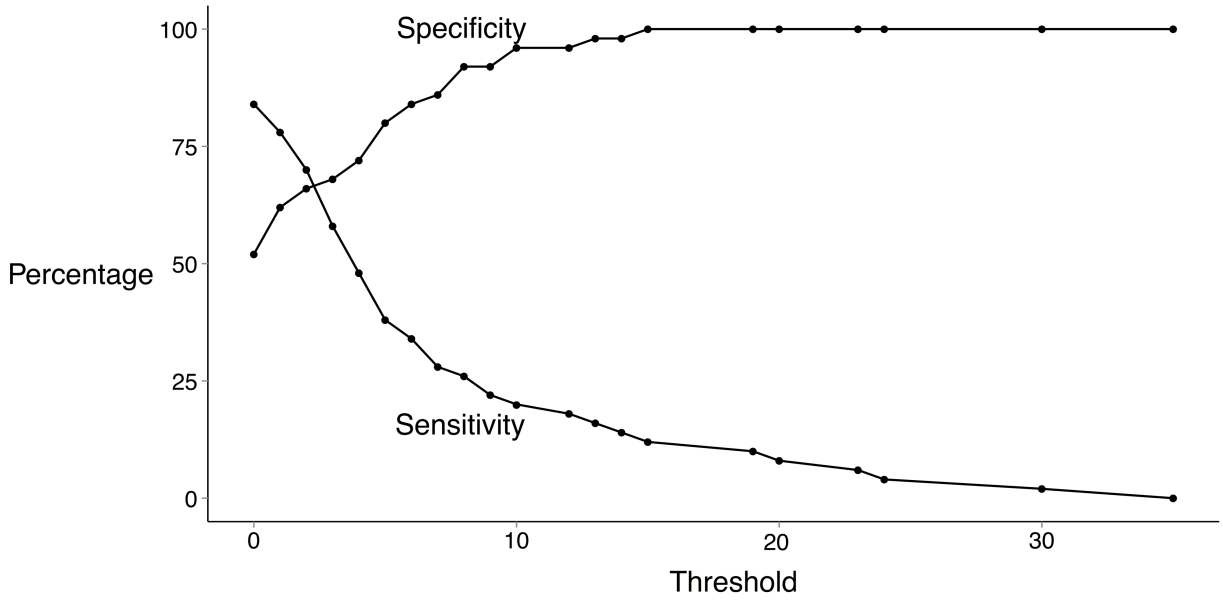


Figure 3. Sensitivity and specificity of the total number of discrepancies in a trial report for retraction.

A reader, unaware of retraction status and applying a cut point of 3 or more discrepancies, would have identified retracted papers with 70% sensitivity and 66% specificity. The usefulness of this in terms of positive predictive value for identification of problems serious enough to cause retraction will depend on prevalence and will therefore vary. We do not suggest trial reports be discounted simply based on reaching a threshold number of discrepancies, but rather that the presence of discrepancies might act as a prompt for the authors to provide the community with access to the raw data, in order to secure trust in the result. In our sample, 92% of unretracted reports had less than ten discrepancies and the majority had none.

Independent identification of discrepancies

Our study design involved three scientists (perhaps simulating reviewers of a manuscript) and one senior scientist (perhaps simulating a final decision-maker on publication of a manuscript). For any discrepancy to be considered valid, all four had to agree that there was no viable explanation present in the trial report. 299 (62%) of the discrepancies were identified by one of the scientists, 78(16%) were independently identified by two scientists and 69(14%) were independently identified by all three scientists. 33(7%) additional discrepancies were noticed by the senior scientist (and subsequently agreed by all others). The time spent by a scientist reading a trial report was available for 269 (90%) of the trial reports. The median time spent by a scientist on a trial report was 23 minutes (interquartile range 11 to 38 minutes).

Consideration of Potential Confounders

We conducted a post-hoc consideration of potential confounders that might mediate an association of discrepancies with retraction status:

- Time (because the rate of retraction of literature may have changed over time)
- Citations (because more frequent citation might signify greater scrutiny) and
- Journal Impact Factor (because retraction has been associated with higher impact factor)

As shown in Table 2, using a zero-inflated negative binomial model, there was no significant association between any of these potential confounders and the number of discrepancies.

Discussion

This study indicates that the presence of discrepancies in a study report should not be assumed to be meaningless. Discrepancies are significantly more common in retracted rather than unretracted articles.

Peer reviewers and other readers may benefit from this knowledge because it is notoriously difficult for them to evaluate the reliability of a trial's findings. It is already known(22,23,24) that the presence of certain features of study design such as blinding, formal enrolment, and automated documentation of results can make a large impact on reported effect size. Our study goes beyond this to suggest that identification of discrepancies, even without particular scientific specialism in the field, might provide an early alarm of unreliability.

Whilst the presence of discrepancies appears sensitive to serious problems within trial reports, it is not specific, in that there are many trial reports in good standing with discrepancies. We do not propose that any particular discrepancy count cut-off should be used as an absolute level for identifying unreliable trial reports. However it might be helpful to identify trial reports where obtaining additional documentation from authors might be important.

Journals could help in additional ways. Providing a post-publication forum for readers to share knowledge of discrepancies is important, because, as our study shows, one reader may spot only a subset of the discrepancies noticed by multiple readers. We believe this finding highlights the difficulty of the task and the likely benefits of crowd-sourcing post-publication examination. In

our study, six retracted and seven unretracted trials had a letter to the editor published or a critical editorial raising concerns. Only one of these letters in each group mentioned any discrepancies.

A journal could plan an automatic "escalation" protocol that would minimize consumption of editorial time. Once it receives a list of discrepancies, it could publish them immediately and request an online supplement of individual patient data from the paper's authors, if such a supplement was not already provided in the original publication. The journal could publish the time in days and hours from request to receipt. In honestly conducted trials with innocent errors for example, honest, simple transcription errors would be identified as such and quickly corrected. Readers might draw their own conclusions if the dataset is delayed or unavailable.(6,25) We propose this approach of requesting the raw data for two reasons. First, if instead the authors were asked to re-run the analyses and/or present an explanation, this could take time to conduct and even more time to achieve agreement between authors; in contrast the raw data can simply be released by the corresponding author, as there should be no debate. Second, such a policy would encourage researchers with nothing to hide to provide the full data as an online supplement in the original publication without waiting for discrepancies to be identified.

Without such protocols and related amendments, journal reviewers and editors must individually find and evaluate the significance of discrepancies. Encouraging instead the many eyes of readers harnesses their crowd-sourced analytical capacity. The readers would know that their observations contribute to science. Journals could respond at an administrative level without

consuming scarce editorial time. Authors would also know that publication would provide genuine scrutiny, not routinely provided or even intended by pre-publication peer review.(26) An alternative mechanism for readers to communicate discrepancies to other readers would be an annotation system such as Pubmed Commons or Pubpeer. This circumvents the system of letters to the editor which is becoming unfit for this purpose because of word count limits and short 6-week statutes of limitations in some journals.(10)

Who would do the work of analyzing the raw data? Meta-analysts are likely to have time and motivation for this, but so would any reader who wanted to find out the correct answer efficiently. The currently practiced approach, which is to write a letter to the editor, is ineffective. For example, even asking about the registered primary endpoint of a trial(27) whose data seemed inexplicably missing from the publication(28) may yield an unrevealing reply.(29) Worse, when the replacement endpoint is discrepant between abstract (and shareholder prospectus(30)) versus individual patient data,(31) the mathematical impossibility can be stonewalled. Worst of all, statistical experts in the field(32) can fail to notice this and instead highlight the queries as being "responsibly rebutted".(33)

Study Limitations

We recognize that even our list of discrepancies (errors) may itself contain errors. Moreover, we have not attempted to establish the mechanism for the discrepancies. We have no way of knowing where each lies on the spectrum from innocent administrative error to intentional fabrication. The strength of our non-judgmental approach is that the presence of discrepancies, rather than any inferred mechanism, is the signal that a trial may be unsound. Author provision

of raw data would allow readers to judge the importance of the identified discrepancies and assist appropriate resolution.

We chose our controls to be the preceding clinical trial in the same journal. Our reasoning was that controlling for journal editorial processes, impact factor,(14) readership and the journal's post-publication policy was the priority. Instead of using a fixed protocol for identifying the control report, it might have been preferable to use individual judgement to select an unretracted report matched for subject matter. However, we considered that attempting to do this would open the study to bias in such selection. Although our study is not able to confidently state whether the observed pattern is changing over time, each control report was very close in time to its counterpart retracted report.

The overall number of discrepancies (of any type) was significantly different between retracted and unretracted trial reports. For every type of discrepancy, the individual count for that type was numerically higher in retracted than unretracted trial reports, and for three of them (arithmetical errors, factual discrepancies and missed p-values) this difference between retracted and unretracted reports was statistically significant. This post-hoc analysis that separated the types of discrepancy did not have power to adequately test each type individually. Whether some types of discrepancy are particularly strong markers of trial report unreliability independently therefore remains uncertain. It is also unknown whether some types of discrepancy imply a mechanism (e.g. fabrication) and should be considered particularly concerning. Alternatively, some readers might be particularly concerned by discrepancies that

have immediate therapeutic implications for patients (e.g. miscalculations of therapeutic effect size or statistical significance).

Another important limitation is that the sample size was constrained by resources because of the time taken to identify, verify and collate the 479 discrepancies in 100 trial reports, and was not statistically based. It would be beneficial if future researchers could address other such trial reports or even reassess the trial reports we assessed.

Misclassified Clinical Trials – A Subgroup Analysis of the Pairs of Reports Meeting a Stricter Definition of "Clinical Trial"

It became apparent during the study that some of the publications identified during the PubMed search would not generally be considered clinical trials of a therapeutic intervention. We performed an additional post-hoc analysis that adopted an aggressive strategy of removing all pairs of PubMed-listed clinical trials where one of the pair was not actually a clinical trial.

This left 68 trial reports in 34 pairs where both would be generally recognised as clinical trials. In Online Supplement 4, we have redrawn all Figures for this subgroup. The pattern we saw in the 100 trial reports classified as clinical trials in Pubmed remained evident in this subgroup of trials with a therapeutic intervention.

There were 335 discrepancies in the 68 trial reports, of which 254 were in the 34 retracted reports and 81 were in the 34 unretracted reports. The overall numbers of discrepancies were 3.1-fold higher for the 34 retracted reports than the 34 unretracted reports. Individual

discrepancy counts remained higher ($p<0.001$) in retracted reports (median 5, IQR 3 to 8.75) than unretracted reports (median 0.5, IQR 0 to 4.75).

32/34 (94%) of retracted reports and 17 (50%) of unretracted reports contained discrepancies. Papers with a discrepancy were significantly more likely to be retracted than those without a discrepancy (OR 16, 95% CI 3.3 to 78, $p<0.001$). The results of the univariable and multivariable zero-inflated negative binomial model in this subgroup showed a trend similar to the larger dataset, albeit, due to a reduced sample size, with wider confidence intervals.

A threshold of 3 or more discrepancies showed sensitivity 76% and specificity 68% for identifying retracted reports in this subgroup.

Conclusion

Discrepancies may be an accessible signal of potential problems in published work. They can be detected without field-specific knowledge. Some discrepancy types are particularly numerous in retracted trials. Recognizing discrepancies may be particularly valuable in clinical trials, where global therapeutic implications mean that many patients' lives and wellbeing may be at stake.

Acknowledgements

Contributors: GDC designed the study, examined the trials, cross-checked the discrepancies, analysed the data, drafted and revised the paper. ANN examined the trials, cross-checked the discrepancies, drafted and revised the paper. MM examined the trials, cross-checked the discrepancies, drafted and revised the manuscript. MJS analysed the data and revised the manuscript. DPF cross-checked the discrepancies, analysed the data, drafted and revised the manuscript. DPF is guarantor.

Funding: This study was not funded by any external organization. All authors are associated with Imperial College London. GDC and MJS are British Heart Foundation clinical research training fellows (FS/12/12/29294 and FS/14/27/30752). DPF is a British Heart Foundation senior clinical research fellow (FS/10/038). MM is supported by an Imperial College Junior Research Fellowship. Neither the institution nor any funder had any role in devising, conducting, analyzing or reporting this study.

Acknowledgement: 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.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in

the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: The complete list of identified discrepancies is shown in Online Supplement 2. These are freely available from the corresponding author in editable form on request. We welcome and will make public any corrections or updates from readers.

Transparency declaration: The guarantor affirms that this manuscript is an honest, accurate and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

Figure Legends

Figure 1. Discrepancy counts. Paired comparison between retracted reports (right) and unretracted control counterparts (left) for all 100 reports studied. Each bar represents the number of discrepancies in one trial. Pairs are ordered by total number of discrepancies in the pair, with those with most discrepancies at the bottom.

Figure 2. Total number of discrepancies by type across all 100 trial reports. Each bar represents the total number of one type of discrepancy in 50 trial reports. White bars are those found in unretracted control reports. Coloured bars are those found in retracted reports.

Figure 3. Sensitivity and specificity of the total number of discrepancies in a trial report exceeding a threshold, and retraction.

Online Supplements

Online Supplement 1

R code used to analyse our data.

Online Supplement 2

Paired comparison between retracted cases (right) and unretracted control counterparts (left). Pairs are ordered by total number of discrepancies in the pair, with those with most discrepancies at the top.

Online Supplement 3

List of discrepancies found in the 100 clinical trial reports.

Online Supplement 4

Figures 1, 2 and 3 redrawn for the post-hoc sub-group of trial reports where we adopted an aggressive strategy of removing all pairs of PubMed-listed clinical trials where one of the pair was not actually a clinical trial of an intervention.

Online Supplement 5

Table 2 for the post-hoc sub-group of trial reports where we adopted an aggressive strategy of removing all pairs of PubMed-listed clinical trials where one of the pair was not actually a clinical trial of an intervention. The upper table shows the univariable analysis. The lower table shows the multivariable analysis.

Online Supplement 6

Dataset to permit re-analysis in R with code available in Online Supplement 1.

References

1 Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature* **483,** 531–533 (2012).

2 Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* **10,** 712 (2011).

3 Ioannidis JPA. Why most published research findings are false. *PLoS Med.* **2,** e124 (2005).

4 Tavare A. Scientific misconduct is worryingly prevalent in the UK, shows BMJ survey. *BMJ* **344,** e377 (2012).

5 Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE* **4,** e5738 (2009).

6 Task Force for Preoperative Cardiac Risk Assessment and Perioperative Cardiac Management in Non-cardiac Surgery. Guidelines for pre-operative cardiac risk assessment and perioperative cardiac management in non-cardiac surgery. *Eur. Heart J.* **30,** 2769–2812 (2009).

7 Erasmus Medical Centre (2012) http://www.erasmusmc.nl/5663/135857/3675250/3706798/Integrity_report_2012-10.pdf?lang=en&lang=en Accessed 21/10/13

8 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* 2013; **100**:456-64.

9 Nowbar AN, Mielewcik M, Karavassilis M, *et al.* Discrepancies in autologous bone marrow stem cell trials and enhancement of ejection fraction (DAMASCENE): weighted regression and meta-analysis. *Br. Med. J.* 2014;**348**:g2688.

10 Francis DP, Mielewczik M, Zargaran D, Cole GD. Autologous bone marrow-derived stem cell therapy in heart disease: Discrepancies and contradictions. *Int. J. Cardiol.* 2013; 168(4): 3381-403.

11 http://www.bmj.com/content/348/bmj.g2688/rapid-responses Accessed 22/11/14

12 Moyé L. DAMASCENE and Meta-Ecological Research: A Bridge Too Far. *Circ Res* 2014;**115**:484–7.

13 Steen RG, Casadevall A, Fang FC. Why has the number of scientific retractions increased? *PLoS ONE* 2013;**8**:e68397.

14 Fang FC, Casadevall A. Retracted science and the retraction index. *Infect Immun* 2011;**79**:3855–9. doi:10.1128/IAI.05661-11

15 Poldermans D, Boersma E, Bax JJ *et al.* Bisoprolol reduces cardiac death and myocardial infarction in high-risk patients as long as 2 years after successful major vascular surgery. *Eur. Heart J.* **22,** 1353–1358 (2001).

16 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.* **12,** 721–729 (2010).

17 Boldt J, Suttner S, Brosch C *et al.* The influence of a balanced volume replacement concept on inflammation, endothelial activation, and kidney integrity in elderly cardiac surgery patients. *Intensive Care Med.* **35,** 462–470 (2009).

18 Quyyumi AA, Waller EK, Murrow J *et al.* CD34(+) cell infusion after ST elevation myocardial infarction is associated with improved perfusion and is dose dependent. *Am. Heart J.* **161,** 98–105 (2011).

19 Bland JM, Altman DG. The odds ratio. BMJ 2000;320:1468

20 Team RDC. R: A Language and Environment for Statistical Computing. Vienna 2006; Available from: http://www.R-project.org/.

21 H. Wickham. ggplot2: elegant graphics for data analysis. Springer New York, 2009

22 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.

23 Shun-Shin MJ, Howard JP, Francis DP. Removing the hype from hypertension. *BMJ* 2014;**348**:g1937.

24 Jabbour RJ, Shun-Shin MJ, Finegold JA, *et al.* Effect of study design on the reported effect of cardiac resynchronization therapy (CRT) on quantitative physiological measures: stratified meta-analysis in narrow-QRS heart failure and implications for planning future studies. *J Am Heart Assoc* 2015;**4**:e000896.

25 Perioperative β blockade: guidelines do not reflect the problems with the evidence from the DECREASE trials. 2014;**349**:g5210.

26 Smith R. Peer review: a flawed process at the heart of science and journals. *J. R. Soc. Med.* **99,** 178–182 (2006).

27 https://www.clinicaltrialsregister.eu/ctr-search/trial/2007-007699-40/BE#A Accessed 04/05/2015

28 Bartunek J, Behfar A, Dolatabadi D, Vanderheyden M, Ostojic M, Dens J, et al. Cardiopoietic stem cell therapy in heart failure: the C-CURE (Cardiopoietic stem Cell therapy in heart failURE) multicenter randomized trial with lineage-specified biologics. *J Am Coll Cardiol.* 2013 ;61(23):2329–38.

29 Bartunek J, Behfar A, Dolatabadi D, *et al.* Reply: The C-CURE Randomized Clinical Trial (Cardiopoietic stem Cell therapy in heart failURE). *J Am Coll Cardiol* 2013;**62**:2454–6.

30 C3 BS. Initial Public Offering of Cardio3 BioSciences [Internet]. Investor Prospectus. 2013 [Accessed 2013 Dec 25]. Available from: http://www.c3bs.com/sites/default/files/financial_presentations/c3bs-prospectus-en.pdf

31 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. Appendix avaialable at http://www.sciencedirect.com/science/article/pii/S073510971305345X#appsec1 (Accessed 04/05/2015)

32 http://projectreporter.nih.gov/Reporter_Viewsh.cfm?sl=13EBCE0A4A84C4D07598B8961CAA4A01A2FFCEB861BF (Accessed 08/12/14)

33 http://www.bmj.com/content/348/bmj.g2688/rr/699167 Accessed 08/05/2015