Article

A Case-Control Comparison of Retracted and Non-Retracted Clinical Trials: Can Retraction Be Predicted?

R. Grant Steen [1],* and Robert M. Hamer [2]

[1] Medical Communications Consultants LLC, 103 Van Doren Place, Chapel Hill, NC 27517, USA

[2] Departments of Psychiatry, Biostatistics, and Psychology, University of North Carolina, Chapel Hill, NC 27517, USA; E-Mail: Hamer@unc.edu

* Author to whom correspondence should be addressed; E-Mail: G_Steen_MediCC@yahoo.com; Tel.: +1-919-942-9299.

Abstract: Does scientific misconduct severe enough to result in retraction disclose itself with warning signs? We test a hypothesis that variables in the results section of randomized clinical trials (RCTs) are associated with retraction, even without access to raw data. We evaluated all English-language RCTs retracted from the PubMed database prior to 2011. Two controls were selected for each case, matching publication journal, volume, issue, and page as closely as possible. Number of authors, subjects enrolled, patients at risk, and patients treated were tallied in cases and controls. Among case RCTs, 17.5% had ≤2 authors, while 6.3% of control RCTs had ≤2 authors. Logistic regression shows that having few authors is associated with retraction ($p < 0.03$), although the number of subjects enrolled, patients at risk, or treated patients is not. However, none of the variables singly, nor all of the variables combined, can reliably predict retraction, perhaps because retraction is such a rare event. Exploratory analysis suggests that retraction rate varies by medical field ($p < 0.001$). Although retraction cannot be predicted on the basis of the variables evaluated, concern is warranted when there are few authors, enrolled subjects, patients at risk, or treated patients. Ironically, these features urge caution in evaluating any RCT, since they identify studies that are statistically weaker.

1. Introduction

Factors that potentially promote research misconduct in clinical research are legion; financial gain, personal fame, scientific hubris and the competitive nature of research funding may all contribute [1]. What should balance these factors is an ethical consideration for the wellbeing of the patient, since fraudulent clinical research may put patients at risk [2–5].

Fabrication and falsification of data has a long history; allegations of misconduct have been made against Ptolemy, Galileo, Newton, Dalton and Mendel [6]. Recently, a claim was made that certain "*warning signs… can be used in most instances to identify… attempts to deceive*" [7], though the editorial in which this claim was made identified no such warning signs. Statistical methods have been devised to detect fraud in clinical questionnaire data [8] and in clinical trial data [6], and statistical comparison of a retracted randomized clinical trial (RCT) to a non-retracted RCT found that statistical features of data in the retracted RCT were so strongly suggestive of data fabrication that other explanations were not plausible [9]. Specifically, differences between treated and control groups at baseline following "randomization" were large enough to suggest that patients had actually not been allocated randomly [9]. However, there are problems with this approach; given the large number of RCTs and the large number of variables measured at baseline, such discrepancies may be frequent. Furthermore, this type of analysis requires access to raw data, which can be precluded by patient confidentiality.

Here, we test a hypothesis that there are warning signs of retraction that may be available in the submitted manuscript, without access to underlying raw data. We ask, are there features of a newly submitted manuscript that should urge caution for editors and referees?

2. Methods

We evaluated every RCT noted as retracted from the PubMed database prior to 2011. PubMed was searched on February 1, 2011, with the limits of "retracted publication, randomized clinical trial, English language." A total of 70 RCTs were identified, all of which were exported from PubMed and saved as a text file (available upon request).

Each retracted paper was read to verify that the research was actually an RCT, defined as a study involving humans prospectively allocated to competing treatments or to treatment and placebo [2]. Papers that met these criteria were further evaluated, using an established analytic approach [3].

Two non-retracted control RCTs were selected to match each retracted case RCT using a method designed to control for standards that might differ between journals or between medical fields. Controls were identified by selecting non-retracted RCTs that had appeared at the same time and in the same journal as the retracted case RCTs; controls were matched to cases by journal name, journal volume and journal issue, with the journal page number matched as closely as possible. This method assumes that control RCTs underwent the same editorial processes as time-matched case RCTs.

Quantitative analysis of retracted RCTs was done using parameters available to any editor prior to external review of a paper. Such data were extracted, as follows [5]:

- **Authors listed**: Total number of named authors under the title;
- **Subjects enrolled**: Total number of patients and healthy controls;

- **Patients at risk**: Patients with an illness (if only healthy subjects were enrolled, this number could be zero)—this is a subset of "Subjects enrolled";
- **Patients treated**: Patients who received a risky intervention (*e.g.*, medication, surgery) in an RCT; blood draw or similarly minor interventions do not count toward this total. Patients who received placebo or standard treatment also do not count. This number could be zero, as it is a subset of "Patients at risk."

Variables were used in a block-matched case-control study, with each block a case and two controls. The primary analysis was a multiple logistic regression to predict retraction status from each of the four variables. As a sensitivity analysis, we used four simple logistic regressions, assessing whether each potential predictor was independently related to retraction status, because correlation among predictors could potentially obscure one or more relationships. Further sensitivity analyses were provided by examining the degree of relationship of retraction status with the variables above using analysis of variance (ANOVA) and a randomization test with a sample of 100,000 observations. We did each analysis with and without a block effect, but, since blocks had essentially no effect, reported results are from analyses without a block effect. All statistical analyses used a statistical package by SAS (Cary, NC, USA).

The number of retracted articles was broken down by field of medicine, as determined by journal title, to determine whether retraction was related to medical field. The total number of RCTs published in a specific field was determined by searching PubMed using the keywords "randomized clinical trials, English language, (field of medicine)." This search strategy enabled us to determine the number of publications per retraction in fields of medicine that had at least one retraction. By searching each year individually, the number of retracted RCTs in, for example, anesthesiology in 2009 could be compared to the total number of published RCTs in anesthesiology in the same year. We collate the numbers of retractions in each field descriptively. A χ^2 analysis was used to evaluate the impression that the number of retractions in anesthesiology was higher than the norm shown in pooled data from all other medical fields that had a retraction.

3. Results and Discussion

3.1. Results

Among the 70 papers identified as retracted RCTs by PubMed, a total of 66 papers (94.3%) were actually evaluated. Four papers were excluded from evaluation for the following reasons: one paper was in an obscure journal that could not be obtained from the medical library of a major American university; one paper was a systematic review misclassified as an RCT; one paper was a psychology study that did not involve patients or treatments; and one study did not involve humans at all. An additional three papers were excluded after further evaluation, because they did not report the results of a clinical trial, but were rather meta-analyses of data from clinical trials. That approximately 10.0% (seven of 70) of studies identified as RCTs by PubMed were misclassified suggests that it can be problematic to rely upon PubMed to identify RCTs.

Cases were matched to controls by journal and by volume in every case, though it was not always possible to match cases and controls by issue number. Of 126 control papers, 116 controls (92.1%) were

selected from the same issue number as the case, implying that control RCTs went through a similar editorial process. Ten controls had to be selected from a different issue than the matched case, but nine of those 10 were selected from the next consecutive issue. One retracted case paper appeared in a journal that published so few RCTs that it was impossible to find two controls in a consecutive issue of the journal. Nevertheless, 99.2% (125/126) of controls were matched to cases within approximately one month of publication in the same journal.

We did all analyses with and without the block effect, and it made no difference in any case. We therefore dropped the block effect for simplicity, and only analyses without the block effect are reported. Table 1 contains descriptive statistics and tests of the effect of each potential predictor in a multiple logistic regression model. By this method, only the number of authors was significantly associated with retraction status ($p < 0.0296$).

Table 1. Comparison of retracted randomized clinical trials (RCTs) (cases) and non-retracted RCTs (controls). "Logistic Regression" predicts retraction from all four predictors one-at-a-time and is significant only for the number of authors. All other analyses do not use a block structure, since this added nothing to the overall predictive ability of the model. "Multiple Logistic Regression" tries to predict retraction from all four predictors in a single equation; the small disagreement between this analysis and the logistic regression suggests that there is some correlation among predictors, such that when all predictors are used together, each adds a small increment to the overall prediction. The log-transformed analysis-of-variance (Log-Transformed ANOVA) uses log-transformed predictors to address likely skew in the data. The "Permutation Test" is used to test the same hypotheses as the ANOVAs, but permutation tests make fewer assumptions. As expected, *p*-values are nearly identical to the ANOVA, indicating that the assumptions of the ANOVA are met. These *p*-values were not corrected for multiple comparisons, because they demonstrate only a mild association that was not useful for prediction. All analyses, except the multiple logistic regression, were sensitivity analyses. Had we corrected *p*-values in the multiple logistic regression, they would not have achieved statistical significance.
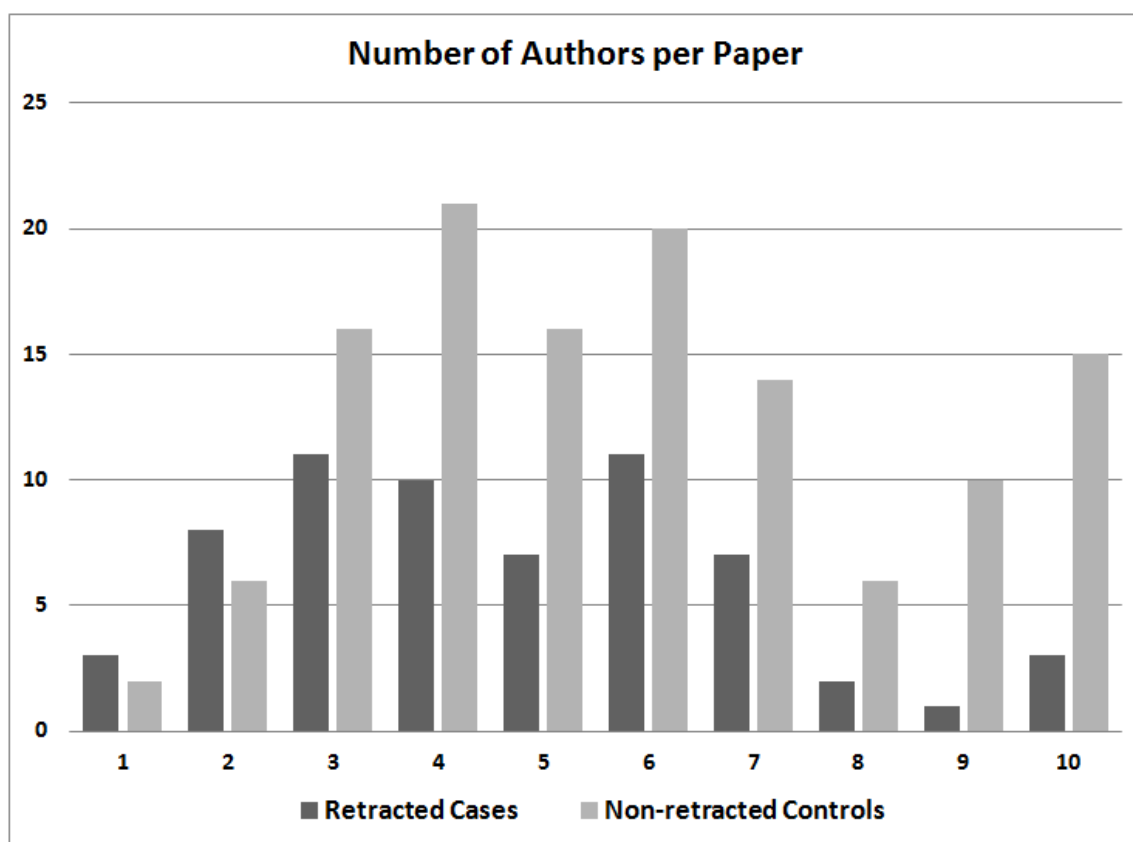
	Retracted Cases ($n = 63$)		Non-Retracted Controls ($n = 126$)		Logistic Regression (1 Variable)	Multiple Logistic regression (All Variables)	Log-Transformed ANOVA (1 Variable)	Permutation Test (100,000 Iterations)
	Mean	**SD**	**Mean**	**SD**	$p \leq$	$p \leq$	$p \leq$	$p \leq$
Number of authors	**5.0**	3.2	**6.7**	5.8	**0.0253**	**0.0296**	**0.0032**	**0.0032**
Subjects enrolled	**216.8**	874.5	**539.6**	2367.6	0.3490	0.6310	0.6061	0.6080
Patients at risk	**198.1**	875.2	**253.5**	1264.5	0.7568	0.6868	0.6816	0.6807
Treated patients	**123.6**	438.4	**167.2**	1216.7	0.7853	0.9957	**0.0047**	**0.0048**

Because the relationship between predictors of retraction and the fact of retraction could have been influenced by other variables in the model, we did four separate logistic regressions to predict the retraction status from each predictor. A small number of RCT authors predicted retraction ($p < 0.0253$) whether the number of authors was the only predictor in the model or other predictors were used.

As an additional test of sensitivity, we used ANOVA to fit models using log-transformation of each of the skewed count variables with retraction status as the predictor. Log(0) was taken as log(0.5), a common manner of handling zeros when using log transformation. Log(number of authors) was significantly related to retraction status ($F(1187) = 8.95$, $p < 0.0032$), as was log(number of treated patients) ($F(1187) = 8.19$; $p < 0.0047$). This implies that, while it is not possible to "significantly" model retraction status as a function of the number of treated patients, retracted papers differ significantly from non-retracted papers in having fewer treated patients. Finally, we assessed these same four variables using permutation tests, with 100,000 samples from the permutation distribution for each of the log-transformed variables. There were significant effects for the log number of authors ($p < 0.0032$) and for the log number of treated patients ($p < 0.0048$).

A plot of the number of authors listed for case and control papers shows that retracted RCTs tend to have fewer named authors than non-retracted control RCTs (Figure 1). It may be noteworthy that 9.5% of retracted RCTs (six of 63) had eight or more authors, whereas 24.6% of control RCTs (31 of 126) had eight or more authors. Conversely, 17.5% of retracted RCTs (11 of 63) had one or two authors, whereas just 6.3% of control RCTs (eight of 126) had so few authors.

Figure 1. The number of authors listed in retracted case randomized clinical trials (RCTs) and non-retracted control RCTs. The category of "10 authors" includes papers with 10 or more authors.

We also determined the number of papers retracted in every field of medicine that had at least one retraction (Table 2). This analysis shows that medical fields differ sharply in the number of retracted RCTs. Anesthesiology had significantly more retracted RCTs than other fields of medicine, even if the

comparison is Bonferroni-corrected for 18 multiple comparisons ($\chi^2 = 94.48$; $p < 0.001$). The impact of "repeat offender" authors was substantial only in anesthesiology, which had 14 retracted RCTs published by a single author (Dr. Scott Reuben); Reuben thus accounted for 63.6% of all RCTs retracted in anesthesiology. Nevertheless, two additional authors also published a retracted fraudulent RCT in anesthesiology. However, if Reuben is deleted from the analysis, anesthesiology is not significantly different from other medical fields.

Table 2. Retracted randomized clinical trials (RCTs) characterized by the field of medicine. "Publications/Retraction" was calculated by dividing "total published" by "total retracted" in each row. Anesthesiology had proportionally more retractions than every other medical field, even when Bonferroni-corrected for 18 possible comparisons between fields ($\chi^2 = 94.48$; $p < 0.001$). Of the 22 retractions in anesthesiology, 14 were first-authored by Scott Reuben.
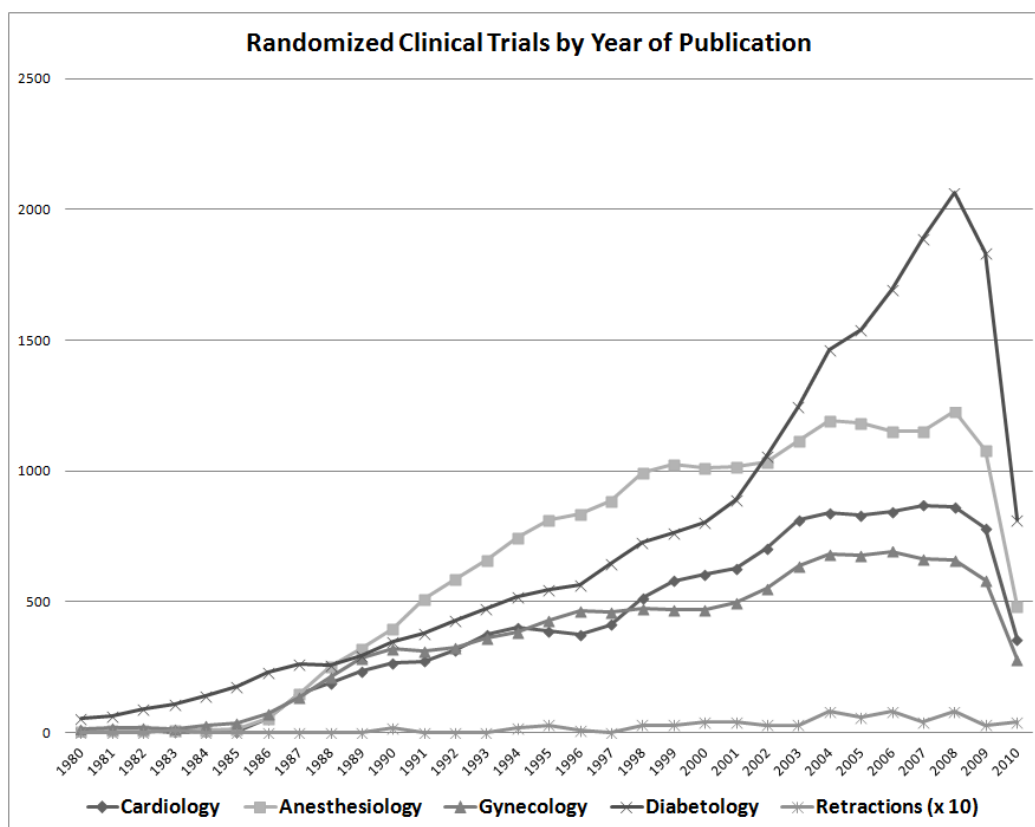
Field	Total Retracted	Total Published	Publications /Retraction	Fraud	Error	Fraudulent Authors
Anesthesiology	22	9881	449	16	6	3
Gynecology	7	6874	982	3	7	3
Surgery	6	66,719	11,120	2	4	2
Oncology	5	6562	1312	2	3	2
Urology/Nephrology	4	103	26	4	0	2
Psychology	4	28,822	7206	4	0	4
Cardiology	3	6259	2086	0	3	0
Pediatrics	2	6832	3416	1	1	1
Rheumatology	1	1604	1604	1	0	1
Diabetology	1	11,055	11,055	1	0	1
Dentistry	1	9054	9054	0	1	0
Emergency medicine	1	1888	1888	0	1	0
Gastroenterology	1	4281	4281	0	1	0
Hematology	1	1913	1913	1	0	1
Orthopedics	1	705	705	0	1	0
Pulmonology	1	3129	3129	1	0	1
Sports medicine	1	607	607	0	1	0
Transplantation	1	6705	6705	1	0	1
Overall (sum or average)	**63**	**172,993**	**2746**	**37**	**29**	**22**

Most retracted RCTs arise from the United States; of the 63 RCTs evaluated (Table 1), 26 (41.3%) were published by first authors whose institution was located in the United States. The next most frequent country from which retracted RCTs arose was Japan, which had five retracted RCTs.

The overall number of retracted RCTs is plotted as a function of year of publication (Figure 2). The number of RCTs published in several fields of medicine is also shown for comparison; diabetology had more RCTs published than did anesthesiology (Figure 2), though anesthesiology had far more RCTs retracted (Table 2). The number of published RCTs has increased progressively until recently, with a commensurate rise in the number of retractions. However, there does not seem to be an increase in the proportion of RCTs retracted, nor is there an obvious relationship between total number of RCTs

retracted and number published within most fields of medicine. Therefore, the high rate of retraction of RCTs in anesthesiology likely cannot be explained by the increase in the overall volume of RCTs in anesthesiology (Table 2).

Figure 2. Randomized clinical trials (RCTs) published per year in several different fields of medicine, together with the number of retracted RCTs (times 10) year-by-year since 1980. The decline in the number of published RCTs after 2008 is striking, especially in diabetology; this may reflect a global downturn in the economy or the lack of funding for clinical research.

3.2. Discussion

Our results suggest that retracted case RCTs have fewer authors and perhaps fewer treated patients than non-retracted control RCTs (Table 1). Yet, differences between cases and controls are not robust enough to predict which papers are likely to be retracted, even when all four variables of interest are combined in a predictive model. Retraction is a very rare event, and given its low prevalence, virtually all positive findings would be false positives, no matter the decision rule. Overall, the greatest association with retraction is having few listed authors (Figure 1), though the field of medicine may also be associated with retraction (Table 2). Until recently, there has been a sharp increase in published RCTs in most fields of medicine (Figure 2), though this increase in publication rate probably cannot explain the large number of retracted RCTs in anesthesiology (Table 2).

Scrupulous evaluation of the data published in retracted RCTs has shown that there can be subtle patterns suggestive of fraud, though evaluating these patterns requires access to raw data. Such patterns can be useful if, for example, a pharmaceutical company is evaluating data from an RCT that has not

been published. Multi-center clinical trials, in particular, offer an opportunity to check the plausibility of submitted clinical data, by comparing data submitted by different centers [6]. In one retracted single-center RCT, the standard deviation of several variables were found to be "unexpectedly" and "unbelievably" low [10]. Furthermore, when *p*-values were recalculated from means and standard deviations in the tables of this paper, recalculated values did not agree with the *p*-values reported by the authors [10]. Such findings argue that RCT publication should require that a locked copy of the dataset, and of the computer programs used to produce results, be deposited in journal archives [10], even if these data are never made available to the public.

Why are retracted case RCTs generally smaller and less ambitious than matched control RCTs? It has been postulated that research fraud represents an effort to obtain the recognition and prominence that a key paper can provide, without actually doing the work [3]. Yet newly-developed statistical methods can detect patterns in data that may be indicative of fraud, with a focus on outliers, inliers, over-dispersion, under-dispersion, hidden correlations, and the lack thereof [6]. Although such methods require access to raw data (and are likely to produce false positives, because retraction is rare), it can be quite difficult to fabricate plausible data [6]. The difficulty of fabricating plausible data confirms that it would be useful to require raw data for an RCT to be filed with the journal of publication.

Clinical trials with a single author are more prone to retraction (Figure 1). This result is consistent with anecdotal reports; Dr. R.K. Chandra published 200 papers, many of which were single-author RCTs, and many of these papers are suspected to be fraudulent [11]. Fraudulent authors also tend to have extraordinarily high research output [10]; such was the case with Dr. R.K. Chandra [11], Dr. Scott Reuben [2] and Dr. Joachim Boldt [12]. This may be because it is less time-consuming to generate a fraudulent RCT than to perform a real one.

The toll from "repeat offender" authors is very damaging and may have skewed our results. Dr. Scott Reuben had 14 retractions among the 63 retracted RCTs examined here, which is 22.2% of all RCTs evaluated (Table 2). Dr. Joachim Boldt had 23 studies retracted in 2011 [12], though only one of his retracted studies is included in this study (the other 22 were retracted after the PubMed search in February, 2011). The Reuben and Boldt cases [13,14] illustrate a common pattern; in-depth investigation inspired by a single retracted paper reveals that a first author has engaged in questionable practices, including other fraudulent papers, sometimes dating back many years [11]. Hence, it is essential that, if research fraud is detected, the entire output of the author in question be carefully examined.

We concur with Trikalinos *et al.*, who concluded that "fraudulent articles are not obviously distinguishable from non-fraudulent ones" [15], although we can add nuance to this conclusion. The Trikalinos study concluded that the number of authors did not differ between fraudulent and non-fraudulent papers [15], yet we found that there was a difference for RCTs, though that difference was too small to be useful as a predictor. The Trikalinos study concluded that there was no difference between research fields in the number of fraudulent papers [15], and we strongly disagree with that conclusion; anesthesiology is the field most prone to retraction (Table 2). Still, without Reuben, anesthesiology would not be more corrupted by misconduct than any other medical field.

What harm might be done to patients by clinical research that is eventually retracted? To evaluate this question, consider just one retracted RCT evaluated here. Some years ago, a very controversial issue in medical oncology was the use of high-dose chemotherapy to treat metastatic breast cancer, with autologous stem cell transplantation after ablation of the patients' immune system [16]. Though several

RCTs had been done to test this therapeutic approach, no study had provided clear-cut evidence that the rigors of treatment were rewarded with improved survival.

Then, a landmark study by Dr. Werner Bezwoda claimed to show that high-dose chemotherapy with stem cell rescue was a useful treatment for breast cancer [17]. Bezwoda had already published many papers in the medical literature; he was a well-established scientist with a commensurate reputation, and his "stem-cell rescue" paper was given credence. In the 1990s, breast cancer became the most common disease for which transplant therapy was given [16]. Yet, it gradually emerged that the key paper [17] was marred by fraud and that at least nine additional papers were also problematic [18]. Eventually, an exhaustive on-site analysis demonstrated unequivocally that the original study could not have been performed as Bezwoda described [16,18].

A great deal of real damage was done to patients by Bezwoda's fraud. According to the National Breast Cancer Coalition, approximately 30,000 breast cancer patients worldwide and 16,000 in the US received high-dose chemotherapy with stem cell rescue prior to when the Bezwoda study was discredited [19]. Many terminally ill breast cancer patients, faced with a choice between standard chemotherapy, which was known to be ineffective in some cases, *versus* an apparently promising therapy, chose high-dose chemotherapy with the attendant side effects. Eventually, even women with early-stage, but high-risk, breast cancer were encouraged to get high-dose chemotherapy, at a cost of $80,000 to $200,000 per treatment [19]. Subsequent clinical trials showed that high-dose treatment with stem cell rescue did little [20] or nothing to extend survival of breast cancer patients [21–24], though such treatment was associated with more frequent adverse events [22], at least in the short term [25]. Hence, a fraudulent RCT did substantial damage to breast cancer patients.

4. Conclusions

Our findings indicate that retraction cannot be predicted on the basis of the *a priori* indicators we tested. However, certain features of RCTs urge caution; having few named authors, few enrolled subjects, few patients at risk or few treated patients are all problematic. Ironically, these features should urge caution in evaluating any RCT, since they indicate that findings are statistically weaker, as well as at greater risk of retraction. We concur that the best overall predictor of the replication of published work is the sample size of subjects enrolled [26], not only because a large sample size is more likely to lead to a replicable finding [27], but also because a large sample size may be harder to fabricate.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Jaffer, U.; Cameron, A.E. Deceit and fraud in medical research. *Int. J. Surg.* **2006**, *4*, 122–126.
2. Steen, R.G. Retractions in the medical literature: Who is responsible for scientific integrity? *AMWA J.* **2011**, *26*, 2–7.
3. Steen, R.G. Retractions in the scientific literature: Do authors deliberately commit research fraud? *J. Med. Ethics* **2011**, *37*, 113–117.

4.   Steen, R.G. Retractions in the scientific literature: Is the incidence of research fraud increasing? *J. Med. Ethics* **2010**, *37*, 249–253.

5.   Steen, R.G. Retractions in the medical literature: How many patients are put at risk by flawed research? *J. Med. Ethics* **2011**, *37*, 688–692.

6.   Buyse, M.; George, S.L.; Evans, S.; Geller, N.L.; Ranstam, J.; Scherrer, B.; LeSaffrez, E.; Murray, G.; Edler, L.; Hutton, J.; *et al.* The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Stat. Med.* **1999**, *18*, 3435–3451.

7.   Harrison, W.T.A.; Simpson, J.; Weil, M. Editorial. *Acta Crystallogr. Sect. E* **2010**, *66*, e1–e2.

8.   Taylor, R.N.; McEntegart, D.J.; Stillman, E.C. Statistical techniques to detect fraud and other data irregularities in clinical questionnaire data. *Drug Inform. J.* **2002**, *36*, 115–125.

9.   Al-Marzouki, S.; Evans, S.; Maeshall, T.; Roberts, I. Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *BMJ* **2005**, *331*, 267–270.

10. White, C. Suspected research fraud: difficulties of getting at the truth. *BMJ* **2005**, *331*, 281–288.

11. Smith, R. Investigating the previous studies of a fraudulent author. *BMJ* **2005**, *331*, 288–291.

12. Marcus, A. More retractions in Boldt case. *Anesthes. News* **2011**, 37. Available online: http://www.anesthesiologynews.com/ViewArticle.aspx?d=Web+Exclusives&d_id=175&i=February+2011&i_id=702&a_id=16700 (accessed on 1 October 2013).

13. Shafer, S.L. Editor's Note: Notices of retraction. *Anesth. Analg.* **2011**, *112*, 1.

14. White, P.F.; Kehlet, H.; Liu, S. Perioperative analgesia: What do we still know? *Anesth. Analg.* **2009**, *108*, 1364–1367.

15. Trikalinos, N.A.; Evangelou, E.; Ioannidis, J.P.A. Falsified papers in high-impact journals were slow to retract and indistinguishable from nonfraudulent papers. *J. Clin. Epidemiol.* **2008**, *61*, 464–470.

16. Weiss, R.B.; Rifkin, R.M.; Stewart, F.M.; Theriault, R.L.; Williams, L.A.; Herman, A.A.; Beveridge, R. A.; High-dose chemotherapy for high-risk primary breast cancer: An on-site review of the Bezwoda study. *Lancet* **2000**, *355*, 999–1003.

17. Bezwoda, W.R.; Seymour, L.; Dansey, R.D. High-dose chemotherapy with hematopoietic rescue as primary treatment for metastatic breast cancer: A randomized trial. *J. Clin. Oncol.* **1995**, *13*, 2483–2489.

18. Weiss, R.B.; Gill, G.G.; Hudis, C.A. An on-site audit of the South African trial of high-dose chemotherapy for metastatic breast cancer and associated publications. *J. Clin. Oncol.* **2001**, *19*, 2771–2777.

19. Editor. Scientific misconduct revealed in high-dose chemotherapy trials. National Council Against Health Fraud Newsletter, **2001**. Available online: http://www.ncahf.org/nl/2001/5–6.html (accessed on 6 September 2013).

20. Rodenhuis, S.; Bontenbal, M.; Beex, L.; Wagstaff, J.; Richel, D.J.; Nooij, M.A.; Voest, E.E.; Hupperets, P.; van Tinteren, H.; Peterse, H.L.; *et al.* High-dose chemotherapy with hematopoietic stem-cell rescue for high-risk breast cancer. *N. Engl. J. Med.* **2003**, *349*, 7–17.

21. Stadtmauer, E.A.; O'Neill, A.; Goldstein, L.J.; Crilley, P.A.; Mangan, K.F.; Ingle, J.N.; Brodsky, I.; Martino, S.; Lazarus, H.M.; Erban, J.K.; *et al.* Conventional-dose chemotherapy compared with high-dose chemotherapy plus autologous hematopoietic stem-cell transplantation for metastatic breast cancer. Philadelphia Bone Marrow Transplant Group. *N. Engl. J. Med.* **2000**, *342*, 1069–1076.

22. Bergh, J.; Wiklund, T.; Erikstein, B.; Lidbrink, E.; Lindman, H.; Malmström, P.; Kellokumpu-Lehtinen, P.; Bengtsson, N.O.; Söderlund, G.; Anker, G.; *et al.* Tailored fluorouracil, epirubicin, and cyclophosphamide compared with marrow-supported high-dose chemotherapy as adjuvant treatment for high-risk breast cancer: A randomised trial. Scandinavian Breast Group 9401 study. *Lancet* **2000**, *356*, 1384–1391.

23. Tallman, M.S.; Gray, R.; Robert, N.J.; LeMaistre, C.F.; Osborne, C.K.; Vaughan, W.P.; Gradishar, W.J.; Pisansky, T.M.; Fetting, J.; Paietta, E.; *et al.* Conventional adjuvant chemotherapy with or without high-dose chemotherapy and autologous stem-cell transplantation in high-risk breast cancer. *N. Engl. J. Med.* **2003**, *349*, 17–26.

24. Hanrahan, E.O.; Broglio, K.; Frye, D.; Buzdar, A.U.; Theriault, R.L.; Valero, V.; Booser, D.J.; Singletary, S.E.; Strom, E.A.; Gajewski, J.L.; *et al.* Randomized trial of high-dose chemotherapy and autologous hematopoietic stem cell support for high-risk primary breast carcinoma: Follow-up at 12 years. *Cancer* **2006**, *106*, 2327–2336.

25. Peppercorn, J.; Herndon, J.; Kornblith, A.B.; Peters, W.; Ahles, T.; Vredenburgh, J.; Schwartz, G.; Shpall, E. Quality of life among patients with Stage II and III breast carcinoma randomized to receive high-dose chemotherapy with autologous bone marrow support or intermediate-dose chemotherapy: Results from Cancer and Leukemia Group B 9066. *Cancer* **2005**, *104*, 1580–1589.

26. Ioannidis, J.P.A. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* **2005**, *294*, 218–228.

27. Ioannidis, J.P.A. Why most published research findings are false. *PLoS Med.* **2005**, *2*, e124.