

# **Publication and Retraction of the Santero, Whitehead, & Ballesteros (2018) Study: A First-Hand Account**

*Neil E. Whitehead<sup>1</sup>*

*Lower Hutt, New Zealand*

This paper details the publication and subsequent unilateral retraction by a bioethics journal of a study showing statistically significant strong effect sizes of combined therapy, strong religiosity, and support groups, on attempts to change sexual orientation by USA men. The study also found very strong and helpful diminution of mental health issues and harm experiences were no worse than those accompanying therapy for other unwanted conditions. The retraction by the journal was nearly a year after publication and was based on a sole negative review of the statistics used, although these had previously been approved after examination in depth by a reviewer from another journal, and one was specifically recommended as valid by the APA. The authors submitted a reply and found the negative review was easily answered, but the editor did not allow herself to be involved in a statistical argument, nor did she blame the authors, but rather attributed culpability to the previous processes in the journal. Gay activism seems not to have been a significant factor, but it will become increasingly necessary for journals to become thoroughly statistically informed.

*Keywords:* sexual orientation change efforts; SOCE; survey; retraction; publication; statistics

---

<sup>1</sup> Neil Evan Whitehead, Ph. D. is a retired scientist, after a career with New Zealand Government research, United Nations, Osaka and Hiroshima universities, and author of more than 150 publications mostly in the hard sciences, but (with Briar Whitehead), author of the book *My Genes Made Me Do It!* His first few degrees, in biochemistry, have gained him an understanding of the degree of biological influences on behaviours. About two dozen publications have been about technical aspects of same-sex attraction (see [www.mygenes.co.nz](http://www.mygenes.co.nz)). He has been involved generally in surveys and their interpretation professionally for 40 years. Correspondence concerning this article could be addressed to him at 54 Redvers Drive, Lower Hutt, New Zealand. 5010. E-mail: [neil@chchquake.co.nz](mailto:neil@chchquake.co.nz)

This study was very similar to the Jones/Yarhouse etc study (Jones & Yarhouse, 2011). In a doctoral thesis survey (P. L. Santero, 2011) started in late 2010 but only published in a journal 7.5 years later (P. Santero, Whitehead, & Ballesteros, 2018), Santero surveyed good and bad experiences of those who had been through therapy for unwanted same-sex attraction. I was recruited to help with statistical aspects, and the following is an account of the saga of getting it published. The tale is mixed and tortuous, and the unexpected outcome almost unique.

I am a research scientist involved in the derivation and interpretation of results in many fields. I have published over 150 papers, in a span of 50 years of research, with lots of use of statistics, but I have to say this has been the most difficult paper to publish of any. My relevant competence level is that I can derive new statistical methods if required, together with their significance limits and necessary software, though this was not needed in the present paper.

Most of my papers have been in the hard sciences—biochemistry, nuclear chemistry and physics, and geological fields, but about two dozen concerned same-sex attraction and closely related issues. The latter were not directly on politically sensitive issues. The Santero paper was, being directly opposed to the recommendations of the American Psychological Association (APA). In their 2009 Task Force Report (American Psychological Association, 2009) they recommended against helping people pursue change who had unwanted same-sex attractions because there was perhaps a risk of harm, and they doubted any change occurred, for example under therapy.

In practice this has resulted in various official bodies banning the therapy and sometimes depriving practitioners of their livelihood, which is an unwarranted

extrapolation from the relatively mild statements in the Task Force Report. In perhaps the most recent case, a South African author has been forbidden even to speak publicly about the subject by their human rights authority. Although in theory all varieties of human rights are equal and have equal protection (e.g. race, sexual orientation, political expression, religious belief), in practice this is far from the case.

An important question in a survey of the effects of therapy is usually about samples. Are the people who have persevered through therapy different from the general population? Can this therapy apply to anyone at all? It turned out that the people in this survey had a religiosity very much higher than the general population. However, they were quite diverse—non-denominational Protestants, Jews, Mormons, a few Catholics, and a few traditional Protestants—no Atheists! The common threads were that they wanted change, and were religiously motivated, had been in therapy, and about half had been in support groups. Probably the most highly motivated did self-select.

The paper showed each set of results in both visual and table form because it seemed there would be one class of people who just wanted the overall message fast, and a second class who might want to check every number carefully!

It was possible to use statistical tests to show same-sex attraction had changed and was not statistically consistent with zero change, but although that is important, it is not enough—one must also show a strong effect. For example, if a drug decreased mortality by 1% it might be a statistically real effect, but it would be far better if it could be shown the drug had a statistically real effect and also reduced mortality by 10%. The percentages are the equivalent of effect sizes in a survey and are important. So one wants to show a significant effect of therapy. This

was certainly possible to calculate although the statistics are less well known than those usually encountered. Furthermore the strength of the effect sizes of the therapy were very comparable to other therapies for unrelated issues. For this group the various therapies work and quite well.

The degree of change was the outcome variable. A number of participants changed a dramatic extent—they said from completely same-sex attraction to completely opposite-sex attraction. Of the whole sample, about two-thirds moved a significant amount towards heterosexual, and the rest mainly did not show any change. A very few actually became more same-sex attracted. However, it was rather remarkable how therapy was found very good, even among those who did not change. One can surmise they had lots of help for other issues and found real fellowship in the support groups.

A rather remarkably wide range of therapeutic techniques had been tried by the respondents. But when asked about whether techniques were helpful or harmful, all techniques received a surprising amount of support, and none were downright harmful. The moral seemed to be that whatever one tried had some positive effect.

Because of accusations that therapy might be damaging to mental health, questions in the survey asked about several issues, such as suicidality, depression, and self-esteem. Again, there were real positive effects with a good effect sizes—in fact, somewhat stronger than for same-sex attraction. So, although before receiving help respondents had an average of three mental health issues, though not severe, they reported lots of help from the therapy and the support groups they attended.

The hypotheses requiring testing, in some cases, caused me much head-scratching to decide the most appropriate statistical approach, not least because there was such a high degree of approval and so few negative

reports. It seemed blatantly obvious there was a profound difference between the positive and negative side, but testing that was tricky—many tests do not react well to having zero people endorsing some survey question! But good tests were ultimately found.

The survey questionnaire had more than 90 questions, and the amount of data generated from 125 participants was large. An outcome was the very strong endorsement of religious issues being the reason for seeking help, with family social pressures being very minor. Distaste for experiences of the gay lifestyle was also a very minor factor.

The major outcome, as found in previous surveys, was that there was real change, little harm, much good, completely opposite to the findings of the APA report, but very close to the findings of Jones and Yarhouse. The question remains: was this a self-selected group? If so the main common factor was religion, and this is remarkably paralleled by the experience of AA, who insist that the help of a Higher Being is essential.

*The Journal of Men's Studies* had previously published a similar report (Karten & Wade, 2010) but this time declined to consider the paper, probably because they did not want to get involved in that controversial area. Another journal not averse to controversy and with high statistical standards gave some excellent comment, both on the statistics and text; the paper was significantly improved, and the statistics accepted. However, they then asked for the names of a couple of competent reviewers outside the highly polarised Western milieu, and this took quite some weeks. (It is common to ask authors for recommendations, and the editors try to find reviewers on the opposite side.) In the interim the journal was sold to another publisher, and a very tight deadline was given for finding the reviewers. They said if it could not be met, the paper would be rejected and treated as a

completely new submission because the new publisher wanted a clean slate. We could not meet the deadline and withdrew the paper.

A submission to the final publication journal was much more positive, and after the usual review process, received an acceptance from the editor, who was told that the statistics had been vetted by another journal. He also asked for statistical review from a reviewer known to him, but this was not forthcoming. The publication was further delayed because of a special theme issue of the journal (these often delay individual papers). This journal was also sold to a new publisher during the review process (again!), but the demands from resubmission were quite minor.

The end result of all this was a 7.5-year gap between survey and publication, and this is very long. In fact only 4% of all papers take more than 5 years (Powell, 2016), so 7.5 years might put this paper in the 2% (?) category. But the results in the paper were so clear that it was well worth persevering. Many scientists complain about the inordinate time it takes to publish a paper. They find that the major conclusions of their papers are not changing in spite of the numerous, protracted, and contradictory reviews, and some are turning to various on-line alternatives.

However, this was far from the end of the saga for the paper. Some months after publication, there was a change of editor, and we received several questions about our qualifications and sources of funding; CVs were asked for. The editor was clearly antsy about the published paper, perhaps through unfamiliarity with the statistics. Nearly a year after publication she announced that the journal was unilaterally withdrawing the paper, on the basis that it had not been adequately statistically reviewed and the review she now received was negative. I mentioned the previous thorough review from another journal, but this did not nullify

the planned retraction. (From inside information it seems gay activism was not a major factor.)

I have not heard of any other such post-publication unilateral retraction except in the case of fraud or plagiarism, neither applicable here. Usually any doubts about a paper's content are addressed in another article in the journal or perhaps a letter to the editor.

The statistical reviewer objected to the diverse methods for therapy employed by the respondents and seemed to think that only very restricted criteria would be adequate (e.g., all subjects should have exactly the same therapy and results should be judged only by interview on carefully established psychometric scales rather than self-reporting). Our reply was that strong effects were seen even under our rather uncontrolled conditions, and this was significant and important. The paper had already said that a main criterion of therapy was whether the client was satisfied, regardless of whether change was small or large, hence self-reports were important. The other objection was that in view of the diversity, a chi-square test was inappropriate. But this is a fundamental misunderstanding of chi-square, which only compares any results obtained with any expected results. No other criteria are needed.

We also mentioned that the APA had recommended another of the less-usual statistical tests we had used. Our statistical reply was apparently not judged adequate by the editor, but there was no reply to our query as to why it was faulty. It rather seems as though the statistical side of this paper was beyond both the editor and reviewer, and our evaluation is that the retraction is totally unjustified.

In retrospect this bioethics journal was faced with strong statistical proof but probably could not really adequately evaluate it because traditionally bioethics arguments hardly involved statistics. However, such expertise will increasingly be needed, and

such journals will have to acquire several expert reviewers.

The published retraction notice (“Retraction notice,” 2019) was curious in another way. It presented the reviewers’ comments and our reply, verbatim, but the editor in the notice said blame should not be ascribed to the authors, implying that the previous editor had not done his job properly.

However, alas, the paper gets into the databases of retractions, with negligible context, and the authors *are* held responsible.

The Linacre Board said they would rather like us to submit a changed manuscript without the statistics, but it was lacking of statistical backing, which was an important criticism of papers by the APA 2009 report, so any such paper would be a non-event. The paper remains on the journal’s website together with its retraction notice.

A further extraordinary facet is the copyright position. As normal we had signed a copyright transfer form, and the unexpected legal position is that the journal, on retraction, has no obligation to return the copyright to us. A journal that wanted to be nasty or protect the academic community from fraud could actually forbid submission for publication anywhere else. The present journal is not so extreme but makes a condition of copyright release that the retraction notice must accompany any submission elsewhere. This would be a major red flag to any journal receiving it and trigger a huge uphill fight. For health reasons I am now not allowed to enter such fights. However, the data remain valid and may be relied on.

Another puzzle is whether data themselves can be copyrighted in the sense of forbidding their use elsewhere. I judge that such a restriction should not apply in academia.

Further publication of similar studies is not likely to be easy.

## References

- American Psychological Association. (2009). *Report of the Task Force on Appropriate Therapeutic Responses to Sexual Orientation*. Washington, DC.
- Jones, S. L., & Yarhouse, M. A. (2011). A longitudinal study of attempted religiously mediated sexual orientation change. *Journal of Sex and Marital Therapy, 37*(5), 404–427. <https://doi.org/10.1080/0092623X.2011.607052>
- Karten, E. Y., & Wade, J. C. (2010). Sexual orientation change efforts in dissatisfied same-sex attracted men. A client perspective. *The Journal of Men’s Studies, 18*, 84–102. <http://dx.doi.org/10.3149/jms.1801.84>
- Powell, K. (2016). Does it take too long to publish research? *Nature, 530*(7589), 148–151. Retrieved from <https://www.nature.com/news/does-it-take-too-long-to-publish-research-1.19320>
- Retraction notice: Effects of therapy on religious men who have unwanted same-sex attraction (2019). *Linacre Quarterly*. <http://dx.doi.org/10.1177/0024363919854842>
- Santero, P. L. (2011). *Change Effects in U.S. Men with Unwanted Same-Sex Attraction after Therapy*. Southern California Seminary.
- Santero, P., Whitehead, N., & Ballesteros, D. (2018). Effects of therapy on religious men who have unwanted same-sex attraction. *Linacre Quarterly, 85*(3), 1–17. <http://dx.doi.org/10.1177/0024363918788559>

